

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 001 090**

51 Int. Cl.:

**C12Q 1/6869** (2008.01)

**G16B 20/00** (2009.01)

**C12Q 1/6809** (2008.01)

**C12Q 1/6883** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.02.2011 E 19197417 (9)**

97 Fecha y número de publicación de la concesión europea: **09.10.2024 EP 3636776**

54 Título: **Métodos para analizar la información de la secuencia de ácido nucleico utilizando el sesgo de GC, opcionalmente para detectar anomalías del ácido nucleico fetal**

30 Prioridad:

**19.02.2010 US 70905710**

**19.03.2010 US 72782410**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**04.03.2025**

73 Titular/es:

**SEQUENOM, INC. (100.00%)**

**3595 John Hopkins Court**

**San Diego, CA 92121, US**

72 Inventor/es:

**LAPIDUS, STANLEY, N.;**

**THOMPSON, JOHN, F.;**

**LIPSON, DORON;**

**MILOS, PATRICE, M.;**

**EFCAVITCH, J., WILLIAM y**

**LETOVSKY, STANLEY**

74 Agente/Representante:

**DEL VALLE VALIENTE, Sonia**

ES 3 001 090 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

**DESCRIPCIÓN**

Métodos para analizar la información de la secuencia de ácido nucleico utilizando el sesgo de GC, opcionalmente para detectar anomalías del ácido nucleico fetal

5

**Campo de la invención**

La invención se refiere generalmente a un método para el recuento cromosómico con efectos reducidos o eliminados del sesgo de GC en la información de secuencia.

10

**Antecedentes**

La aneuploidía fetal (p. ej., el síndrome de Down, el síndrome de Edward y el síndrome de Patau) y otras anomalías cromosómicas afectan a 9 de cada 1000 nacidos vivos (Cunningham y col. en Williams Obstetrics, McGraw-Hill, Nueva York, pág. 942, 2002). Las anomalías cromosómicas generalmente se diagnostican mediante el cariotipo de células fetales obtenidas mediante procedimientos invasivos, como la toma de muestras de vellosidades coriónicas o la amniocentesis. Estos procedimientos se asocian con riesgos potencialmente importantes tanto para el feto como para la madre. Se dispone de exámenes de detección no invasivos con marcadores séricos maternos o ecografías, pero su fiabilidad es limitada (Fan y col., PNAS, 105(42): 16266-16271, 2008).

15

20

Desde el descubrimiento de células fetales intactas en la sangre materna, ha habido un gran interés en tratar de utilizar esas células como una ventana de diagnóstico en la genética fetal (Fan y col., PNAS, 105(42): 16266-16271, 2008). El descubrimiento de que existen ciertas cantidades (entre aproximadamente el 3 % y aproximadamente el 6 %) de ácidos nucleicos fetales libres de células en la circulación materna ha llevado al desarrollo de pruebas genéticas prenatales basadas en PCR no invasivas para una variedad de rasgos. Un problema con esas pruebas es que los ensayos basados en la PCR intercambian la sensibilidad por la especificidad, lo que dificulta la identificación de mutaciones particulares. Además, debido a la naturaleza estocástica de la PCR, frecuentemente se pasa por alto una población de moléculas que está presente en una pequeña cantidad en la muestra, como el ácido nucleico fetal en una muestra de un tejido o fluido corporal materno. De hecho, si el ácido nucleico raro no se amplifica en las primeras rondas de amplificación, es cada vez más improbable que el evento raro se detecte alguna vez.

25

30

Además, también existe la posibilidad de que el ácido nucleico fetal en una muestra materna se degrade y no pueda modificarse para la amplificación por PCR debido al pequeño tamaño del ácido nucleico.

35

Existe la necesidad de métodos que puedan detectar de forma no invasiva los ácidos nucleicos fetales y diagnosticar anomalías fetales.

Alkan y col. (Nat Genet, 41(10): 1061-1067, 2009) exponen un algoritmo para mapear las lecturas de secuencias de última generación. Los autores utilizaron una corrección estadística para corregir los sesgos del GC.

40

Chu y col. (Bioinformatics 25(10): 1244-50, 2009) describe un método de secuenciación del genoma completo para el diagnóstico de aneuploidía. Los autores desarrollaron un modelo estadístico para corregir automáticamente el sesgo de GC en los datos de secuenciación del genoma completo.

45

La patente WO 2010/033578 A2 describe un método para analizar una muestra materna que se recoge de forma no invasiva. El método implica generar un gran número de lecturas de secuencias cortas de ADN materno y fetal. El método implica además etapas para corregir el sesgo de densidad de marcadores de secuencia.

**Resumen**

50

La invención se expone en las reivindicaciones adjuntas. Los métodos de la invención aprovechan las tecnologías de secuenciación, particularmente las tecnologías de secuenciación por síntesis de moléculas individuales, para detectar el ácido nucleico fetal en los tejidos o fluidos corporales maternos. Los métodos de la invención son altamente sensibles y permiten la detección de la pequeña población de ácidos nucleicos fetales en una muestra materna, generalmente sin la necesidad de amplificar el ácido nucleico en la muestra.

55

Los métodos de la invención implican secuenciar el ácido nucleico obtenido de una muestra materna y distinguir entre el ácido nucleico materno y fetal. La distinción entre el ácido nucleico materno y fetal identifica el ácido nucleico fetal, que permite por lo tanto la determinación de anomalías con respecto a la variación de la secuencia. Dichas anomalías pueden determinarse como polimorfismos de un solo nucleótido, motivos variantes, inversiones, deleciones, adiciones o cualquier otro reordenamiento o anomalía de los ácidos nucleicos.

60

Los métodos de la invención también se usan para determinar la presencia de ácido nucleico fetal en una muestra materna identificando el ácido nucleico que es exclusivo del feto. Por ejemplo, se pueden buscar diferencias entre la secuencia obtenida y la secuencia de referencia materna; o puede implicar la identificación del material cromosómico

65

Y en la muestra. La muestra materna puede ser un tejido o un fluido corporal. En realizaciones particulares, el fluido corporal es sangre materna, plasma sanguíneo materno o suero materno.

5 La invención también proporciona una forma de confirmar la presencia de ácido nucleico fetal en una muestra materna, por ejemplo, buscando secuencias o variantes únicas.

10 La reacción de secuenciación puede ser cualquier reacción de secuenciación. En realizaciones particulares, la reacción de secuenciación es una reacción de secuenciación de una sola molécula. La secuenciación de una sola molécula se muestra, por ejemplo, en Lapidus y col. (patente US-7.169.560), Lapidus y col. (solicitud de la patente US-2009/0191565), Quake y col. (patente US-6.818.395), Harris (patente US-7.282.337), Quake y col. (solicitud de la patente US-2002/0164629), y Braslavsky, y col., PNAS (EE. UU.), 100: 3960-3964 (2003).

15 En resumen, en algunas implementaciones, un ácido nucleico monocatenario (p. ej., ADN o ADNc) se hibrida con oligonucleótidos unidos a una superficie de una célula de flujo. Los oligonucleótidos pueden unirse covalentemente a la superficie o pueden emplearse diversas uniones distintas de la unión covalente, tal como conocen los expertos en la materia. Además, la unión puede ser indirecta, p. ej., a través de las polimerasas de la invención unidas directa o indirectamente a la superficie. La superficie puede ser plana o de otro tipo, y/o puede ser porosa o no porosa, o cualquier otro tipo de superficie que los expertos en la técnica sepan que es adecuada para la unión. A continuación, el ácido nucleico se secuencia mediante imágenes o detectando de otro modo la adición mediada por la polimerasa de nucleótidos marcados con fluorescencia incorporados en el oligonucleótido de la superficie de la cadena en crecimiento, con una resolución de una sola molécula. En ciertas realizaciones, los nucleótidos usados en la reacción de secuenciación no son nucleótidos que terminan la cadena.

20 Debido a que el cromosoma Y solo estará presente si el ácido nucleico fetal es de un varón, los métodos de la invención pueden incluir además realizar un ensayo cuantitativo en las secuencias obtenidas para detectar la presencia de ácido nucleico fetal si el cromosoma Y no se detecta en la muestra. Dichos ensayos cuantitativos incluyen el análisis del número de copias, la llamada de alelos dispersos, la resecuenciación dirigida y el análisis de puntos de quiebre.

25 La capacidad de detectar el ácido nucleico fetal en una muestra materna permite el desarrollo de un ensayo de diagnóstico no invasivo para evaluar si el feto tiene una anomalía. Por lo tanto, otro aspecto de la invención proporciona métodos no invasivos para determinar si un feto tiene una anomalía. Los métodos de la invención pueden implicar obtener una muestra que incluya ácidos nucleicos maternos y fetales, realizar una reacción de secuenciación en la muestra para obtener información de secuencia sobre los ácidos nucleicos de la muestra, comparar la información de secuencia obtenida con la información de secuencia de un genoma de referencia, determinando de este modo si el feto tiene una anomalía, detectar la presencia de al menos una parte de un cromosoma Y en la muestra y distinguir los falsos negativos de los verdaderos negativos si el cromosoma Y no se detecta en la muestra.

30 Un aspecto importante de un ensayo de diagnóstico es la capacidad del ensayo para distinguir entre los falsos negativos (no se detecta el ácido nucleico fetal cuando de hecho está presente) y los negativos verdaderos (detección del ácido nucleico de un feto sano). Los métodos de la descripción proporcionan esta capacidad. Si el cromosoma Y se detecta en la muestra materna, los métodos de la descripción aseguran que el ensayo funciona correctamente, porque el cromosoma Y está asociado solo con los hombres y estará presente en una muestra materna solo si el ácido nucleico fetal masculino está presente en la muestra. Algunos métodos de la descripción proporcionan un análisis cuantitativo o cualitativo adicional para distinguir entre falsos negativos y verdaderos negativos, independientemente de la capacidad de detectar el cromosoma Y, particularmente para muestras que incluyen ácidos nucleicos normales de un feto femenino. Tal análisis cuantitativo adicional puede incluir el análisis del número de copias, la llamada de alelos dispersos, la resecuenciación dirigida y el análisis de puntos de corte.

35 Otro aspecto de la descripción proporciona métodos para determinar si un feto tiene una anomalía, incluida la obtención de una muestra materna que comprende ácidos nucleicos maternos y fetales; unir marcadores únicos a los ácidos nucleicos de la muestra, en las que cada marcador está asociado con un cromosoma diferente; realizar una reacción de secuenciación en los ácidos nucleicos marcados para obtener secuencias marcadas; y determinar si el feto tiene una anomalía cuantificando las secuencias marcadas. En ciertas realizaciones, los marcadores incluyen secuencias de ácido nucleico únicas.

55 **Breve descripción de los dibujos**

60 La Figura 1 es un histograma que muestra la diferencia entre un individuo (“yo”) y dos miembros de la familia (“familia”) que representa una comparación de un conjunto de variantes conocidas de un solo nucleótido entre las tres muestras.

65 La Figura 2 es una tabla que muestra las lecturas de secuencias de ADN de HapMap derivadas de la secuenciación de una sola molécula y alineadas de forma única con un genoma humano de referencia. Cada columna representa datos de un único canal secuenciador HELISCOPE (aparato de secuenciación de una sola molécula, Helicos BioSciences Corporation).

La Figura 3 es una tabla que muestra las lecturas cromosómicas normalizadas por muestra. Los recuentos cromosómicos individuales se dividieron por los recuentos autosómicos totales.

5 La Figura 4 es una tabla que muestra los recuentos normalizados por cromosoma. La fracción promedio de lecturas alineadas con cada cromosoma en todas las muestras.

La Figura 5 es una representación gráfica de los recuentos cromosómicos cuantitativos.

10 La Figura 6 es un gráfico que muestra una muestra en la que los recuentos cromosómicos están sesgados por el sesgo de GC.

15 La Figura 7 es un gráfico que muestra los bins genómicos representados como una función del contenido de GC en el bin. En la Figura 7, la muestra superior muestra una correlación positiva con el contenido de GC, y la muestra inferior muestra una correlación negativa con el contenido de GC.

El panel A de la Figura 8 es un gráfico que muestra la selección de ciertos bins genómicos con un contenido de GC dado para el análisis. El panel B de la Figura 8 muestra la información de la secuencia antes de la corrección del sesgo de GC. El panel C de la Figura 8 muestra la información de la secuencia después de la corrección del sesgo de GC.

20 Los paneles A y B de la Figura 9 muestran información de secuencia antes de la corrección del sesgo de GC.

Los paneles C y D de la Figura 9 muestran información de secuencia después de la corrección del sesgo de GC.

25 La Figura 10 muestra los resultados del análisis de la información de secuencia.

### **Descripción detallada**

30 Los métodos de la invención usan reacciones de secuenciación para detectar la presencia de ácido nucleico fetal en una muestra materna. Los métodos de la invención también usan reacciones de secuenciación para analizar la sangre materna para detectar una afección genética, en las que se analiza el ácido nucleico mixto fetal y materno en la sangre materna para distinguir una mutación fetal o una anomalía genética del fondo del ácido nucleico materno.

35 El ácido nucleico fetal incluye tanto el ADN fetal como el ARN fetal. Como se describe en Ng y col., el ARNm de origen placentario es fácilmente detectable en el plasma materno, Proc. Nat. Acad. Sci. 100(8): 4748-4753 (2003).

### **Muestras**

40 Los métodos de la invención implican obtener una muestra, p. ej., un tejido o fluido corporal, que se sospecha que incluye ácidos nucleicos maternos y fetales. Dichas muestras pueden incluir saliva, orina, lágrimas, secreción vaginal, líquido amniótico, líquido mamario, leche materna, sudor o tejido. En ciertas realizaciones, esta muestra se extrae de sangre materna y el ADN circulante se encuentra en el plasma sanguíneo, en lugar de en las células. Una muestra preferida es la sangre venosa periférica materna.

45 En ciertas realizaciones, se extraen aproximadamente 10-20 ml de sangre. Esa cantidad de sangre permite obtener al menos aproximadamente 10.000 equivalentes genómicos del ácido nucleico total (el tamaño de la muestra se basa en una estimación de la presencia de ácido nucleico fetal en aproximadamente 25 equivalentes genómicos/ml de plasma materno al principio del embarazo, y una concentración de ácido nucleico fetal de aproximadamente el 3,4 % del ácido nucleico plasmático total). Sin embargo, se puede extraer menos sangre para un análisis genético cuando se requiere menos significación estadística, o la muestra de ácido nucleico se enriquece en ácido nucleico fetal.

50 Debido a que la cantidad de ácido nucleico fetal en una muestra materna generalmente aumenta a medida que avanza el embarazo, es posible que se requiera menos muestra a medida que avanza el embarazo para obtener la misma cantidad o una cantidad similar de ácido nucleico fetal de una muestra.

### **Enriquecimiento**

55 En ciertas realizaciones, la muestra (p. ej., sangre, plasma o suero) puede enriquecerse opcionalmente en ácido nucleico fetal mediante métodos conocidos, tales como el fraccionamiento por tamaño para seleccionar fragmentos de ADN de menos de aproximadamente 300 pares de bases. Alternativamente, puede excluirse el ADN materno, que tiende a ser mayor de aproximadamente 500 pares de bases.

60 En ciertas realizaciones, la sangre materna puede procesarse para enriquecer la concentración de ADN fetal en el ADN total, como se describe en Li y col., J. Amer. Med. Assoc. 293:843-849, 2005). En resumen, el ADN circulatorio se extrae de 5 ml a 10 ml de plasma materno utilizando tecnología de columna comercial (kit de purificación de ADN Roche High Pure Template; Roche, Basilea, Suiza) en combinación con una bomba de vacío. Tras la extracción, el ADN se separa mediante electroforesis en gel de agarosa (1 %) (Invitrogen, Basilea, Suiza) y se extirpa

cuidadosamente la fracción de gel que contiene ADN circulatorio con un tamaño de aproximadamente 300 pb. El ADN se extrae de este corte de gel utilizando un kit de extracción (kit de extracción en gel QIAEX II); Qiagen, Basilea, Suiza) y se eluyó en un volumen final de 40 µl de ácido trihidroclórico 10 mM estéril, pH 8,0 (Roche).

5 El ADN se puede concentrar mediante métodos conocidos, que incluyen la centrifugación y varios inhibidores enzimáticos. El ADN se une a una membrana selectiva (p. ej., sílice) para separarlo de los contaminantes. El ADN se enriquece preferiblemente en fragmentos que circulan en el plasma, que tienen una longitud inferior a 1000 pares de bases, generalmente inferiores a 300 pb. Esta selección del tamaño se realiza en un medio de separación del tamaño del ADN, tal como un gel electroforético o un material de cromatografía. Un material de este tipo se describe en Huber y col. (Nucleic Acids Res. 21(5): 1061-1066, 1993), cromatografía de filtración de gel, TSKgel, como se describe en Kato y col., (J. Biochem, 95(1):83-86, 1984).

Además, el enriquecimiento puede lograrse mediante la supresión de ciertos alelos mediante el uso de ácidos nucleicos péptídicos (PNA), que se unen a sus secuencias diana complementarias, pero no se amplifican.

15 La extracción de ARN plasmático se describe en Enders y col. (Clinical Chemistry 49:727-731, 2003). Como se describe allí, el plasma recogido después de las etapas de centrifugación se mezcla con el reactivo Trizol LS (Invitrogen) y cloroformo. La mezcla se centrifuga y la capa acuosa se transfiere a tubos nuevos. Se añade etanol a la capa acuosa. La mezcla se aplica después a una minicolumna RNeasy (Qiagen) y se procesa según las recomendaciones del fabricante.

Otra etapa de enriquecimiento puede ser tratar la muestra de sangre con formaldehído, como se describe en Dhallan y col. (J. Am. Med. Soc. 291(9): 1114-1119, marzo de 2004; y solicitud de la patente US-20040137470). Dhallan y col. (La solicitud de la patente US-20040137470) describe un procedimiento de enriquecimiento del ADN fetal, en donde se recoge sangre en tubos Vacuette con EDTA de 9 ml (número de catálogo NC9897284) y se añaden a cada tubo 0,225 ml de solución tamponada neutra al 10 % que contiene formaldehído (4 % p/v), y cada tubo se invierte suavemente. Los tubos se almacenan a 4 °C hasta que estén listos para su procesamiento.

Los agentes que impiden la lisis celular o estabilizan las membranas celulares se pueden añadir a los tubos, incluidos, entre otros, formaldehído y derivados de formaldehído, formalina, glutaraldehído y derivados del glutaraldehído, reticulantes, reticulantes reactivos con aminos primarias, reticulantes reactivos con sulfhidrilo, adición de sulfhidrilo o reducción de disulfuro, reticulantes reactivos con carbohidratos, reticulantes reactivos con carboxilo, reticulantes fotoactivos, reticulantes escindibles, etc. Se puede añadir cualquier concentración de agente que estabilice las membranas celulares o impida la lisis celular. En ciertas realizaciones, el agente que estabiliza las membranas celulares o impide la lisis celular se añade a una concentración que no impide ni dificulta las reacciones posteriores.

Las técnicas de citometría de flujo también se pueden usar para enriquecer las células fetales (Herzenberg y col., PNAS 76:1453-1455, 1979; Bianchi y col., PNAS 87:3279-3283, 1990; Bruch y col., Prenatal Diagnosis 11:787-798, 1991). Saunders y col. (patente US-5.432.054) también describen una técnica para la separación de glóbulos rojos nucleados fetales, utilizando un tubo que tiene una parte superior ancha y una parte inferior capilar estrecha hecha de polietileno. La centrifugación mediante un programa de velocidad variable da como resultado un apilamiento de glóbulos rojos en el capilar con respecto a la densidad de las moléculas. La fracción de densidad que contiene glóbulos rojos de baja densidad, incluidos los glóbulos rojos fetales, se recupera y luego se hemoliza diferencialmente para destruir preferiblemente los glóbulos rojos maternos. Se utiliza un gradiente de densidad en un medio hipertónico para separar los glóbulos rojos, ahora enriquecidos en glóbulos rojos fetales, de los linfocitos y las células maternas rotas. El uso de una solución hipertónica reduce el tamaño de los glóbulos rojos, lo que aumenta su densidad y facilita la purificación de los linfocitos más densos. Una vez aisladas las células fetales, el ADN fetal se puede purificar usando técnicas estándar en la materia.

Además, se puede añadir un agente que estabiliza las membranas celulares a la sangre materna para reducir la lisis de las células maternas, incluidos, entre otros, aldehídos, urea-formaldehído, fenol formaldehído, DMAE (dimetilaminoetanol), colesterol, derivados del colesterol, altas concentraciones de magnesio, vitamina E y derivados de la vitamina E, calcio, gluconato de calcio, taurina, niacina, derivados de hidroxilamina, bimocromol, asa, astaxantina, glucosa, amitriptilina, isómero A, fenilacetato tetral de hopano-isómero B, fenilacetato de hopano-tetral, citicolina, inositol, vitamina B, complejo de vitamina B, hemisuccinato de colesterol, sorbitol, calcio, coenzima Q, ubiquinona, vitamina K, complejo de vitamina K, menaquinona, zonegran, zinc, extracto de ginkgo biloba, difenilhidantoína, perftoran, polivinilpirrolidona, fosfatidilserina, tegretol, PABA, cromglicato disódico, nedocromil sodio, feniloína, citrato de zinc, mexitil, dilantina, hialuronato de sodio o polaxámero 188.

Un ejemplo de protocolo para usar este agente es el siguiente: La sangre se almacena a 4 °C hasta su procesamiento. Los tubos se hacen girar a 1000 rpm durante diez minutos en una centrifuga con la potencia de frenado ajustada a cero. Los tubos se centrifugan una segunda vez a 1000 rpm durante diez minutos. El sobrenadante (el plasma) de cada muestra se transfiere a un tubo nuevo y se hace girar a 3000 rpm durante diez minutos con el freno puesto a cero. El sobrenadante se transfiere a un tubo nuevo y se almacena a -80 °C. Aproximadamente dos mililitros de la "capa leucocitaria", que contiene células maternas, se colocan en un tubo separado y se almacenan a -80 °C.

El ADN genómico se puede aislar del plasma utilizando el kit Qiagen Midi para la purificación del ADN de las células sanguíneas, siguiendo las instrucciones del fabricante (kit QiAmp DNA Blood Midi, número de catálogo 51183). El ADN se eluye en 100 µl de agua destilada. El kit Midi de Qiagen también se usa para aislar el ADN de las células maternas contenidas en la "capa leucociana".

#### Extracción

El ácido nucleico se extrae de la muestra según métodos conocidos en la técnica. Véase, por ejemplo, Maniatis y col., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., págs. 280-281, 1982.

#### Determinación de la presencia de ácido nucleico fetal masculino en una muestra materna

El ácido nucleico de la muestra se analiza luego usando una reacción de secuenciación para detectar la presencia de al menos una porción de un cromosoma Y en la muestra. Por ejemplo, Bianchi y col. (PNAS USA, 87:3279-3283, 1990) informa de una secuencia de 222 pares de bases que está presente solo en el brazo corto del cromosoma Y. Lo y col. (Lancet, 350:485-487, 1997), Lo, y col., (Am J Hum Genet, 62(4):768, 1998) y Smid y col. (Clin Chem, 45:1570-1572, 1999) cada uno informa de diferentes secuencias cromosómicas Y derivadas de fetos masculinos. Si el cromosoma Y se detecta en la muestra materna, los métodos de la invención aseguran que la muestra incluya ácido nucleico fetal, porque el cromosoma Y está asociado solo con hombres y estará presente en una muestra materna solo si el ácido nucleico fetal masculino está presente en la muestra.

En ciertas realizaciones, el método de secuenciación es un método de secuenciación de una sola molécula por síntesis. La secuenciación de una sola molécula se muestra, por ejemplo, en Lapidus y col. (patente US-7.169.560), Lapidus y col. (solicitud de la patente US-2009/0191565), Quake y col. (patente US-6.818.395), Harris (patente US-7.282.337), Quake y col. (solicitud de la patente US-2002/0164629), y Braslavsky, y col., PNAS (EE. UU.), 100: 3960-3964 (2003).

En resumen, un ácido nucleico monocatenario (p. ej., ADN o ADNc) se hibrida con oligonucleótidos unidos a una superficie de una célula de flujo. Los oligonucleótidos pueden unirse covalentemente a la superficie o pueden emplearse diversas uniones distintas de la unión covalente, tal como conocen los expertos en la materia. Además, la unión puede ser indirecta, p. ej., mediante una polimerasa unida directa o indirectamente a la superficie. La superficie puede ser plana o de otro tipo, y/o puede ser porosa o no porosa, o cualquier otro tipo de superficie que los expertos en la técnica sepan que es adecuada para la unión. A continuación, el ácido nucleico se secuencia mediante la obtención de imágenes de la adición mediada por la polimerasa de nucleótidos marcados con fluorescencia incorporados en el oligonucleótido de la superficie de la cadena en crecimiento, con una resolución de una sola molécula. En ciertas realizaciones, los nucleótidos usados en la reacción de secuenciación no son nucleótidos que terminan la cadena. Las siguientes secciones analizan las consideraciones generales para la secuenciación de ácidos nucleicos, por ejemplo, las polimerasas útiles en la secuenciación por síntesis, la elección de superficies, las condiciones de reacción, la detección y el análisis de señales.

#### Nucleótidos

Los nucleótidos útiles en la invención incluyen cualquier nucleótido o análogo de nucleótido, ya sea de origen natural o sintético. Por ejemplo, los nucleótidos preferidos incluyen ésteres de fosfato de desoxiadenosina, desoxicitidina, desoxiguanosina, desoxitimidina, adenosina, citidina, guanosina y uridina. Otros nucleótidos útiles en la invención comprenden una base de adenina, citosina, guanina, timina, xantina o hipoxantina; 5-bromouracilo, 2-aminopurina, desoxiinosina o citosina metilada, tal como 5-metilcitosina y N4-metoxidesoxicitosina. También se incluyen bases de polinucleótidos miméticos, tales como ácidos nucleicos metilados, p. ej., 2'-O- metRNA, ácidos nucleicos peptídicos, ácidos nucleicos peptídicos modificados, ácidos nucleicos bloqueados y cualquier otro resto estructural que pueda actuar sustancialmente como un nucleótido o una base, p. ej., exhibiendo complementariedad de bases con una o más bases que se encuentran en el ADN o el ARN y/o que puedan incorporarse como bases complementarias, e incluyen análogos que terminan la cadena. Un nucleótido corresponde a una especie de nucleótido específica si comparten complementariedad de bases con respecto a al menos una base.

Los nucleótidos para la secuenciación de ácidos nucleicos según la invención incluyen preferiblemente un marcador detectable que es detectable directa o indirectamente. Los marcadores preferidos incluyen marcadores detectables ópticamente, tales como marcadores fluorescentes. Los ejemplos de marcadores fluorescentes incluyen, aunque no de forma limitativa, colorantes Atto, ácido 4-acetamido-4'-isotiocianatoestilbeno-2,2'disulfónico; acridina y derivados: acridina, isotiocianato de acridina; ácido 5-(2'-aminoetil)aminonaftalen-1-sulfónico (EDANS); 3,5-disulfonato de 4-amino-N-[3-vinilsulfonil]fenil]naftalimida; N-(4-anilino-1-naftil)maleimida; antranilamida; BODIPY; amarillo brillante; cumarina y derivados; cumarina, 7-amino-4-metilcumarina (AMC, Cumarina 120), 7-amino-4-trifluorometilcouluarina (Coumaran 151); tintes de cianina; cianosina; 4',6-diaminidino-2-fenilindol (DAPI); dibromopirogalol-sulfonaftaleína de 5 pies y 5 pulgadas (rojo de bromopirogalol); 7-dietilamino-3-(4'-isotiocianatofenil)-4-metilcumarina; pentaacetato de dietilentriamina; ácido 4,4'-diisotiocianatodihidro-estilbeno-2,2'-disulfónico; ácido 4,4'-diisotiocianatoestilbeno-2,2'-disulfónico; cloruro de 5-[dimetilamino]naftaleno-1-sulfonilo (DNS, cloruro de dansilo); 4-dimetilaminofenilazofenil-4'-isotiocianato (DABITC); eosina y derivados; eosina, isotiocianato de eosina, eritrosina y derivados; eritrosina B,

eritrosina, isotiocianato; etidio; fluoresceína y derivados; 5-carboxifluoresceína (FAM), 5-(4,6-diclorotriazin-2-il)aminofluoresceína (DTAF), 2',7'-dimetoxi-4'5'-dicloro-6-carboxifluoresceína, fluoresceína, isotiocianato de fluoresceína, QFITC, (XRITC); fluorescamina; IR144; IR1446; isotiocianato verde de malaquita; 4-metilumbeliferona orto-cresoltaleína; nitrotirosina; pararosanilina; rojo de fenol; B-ficoeritrina; o-ftaldialdehído; pireno y derivados: pireno, butirato de pireno, succinimidil-1-pireno; puntos cuánticos de butirato; rojo reactivo 4 (Cibacron.TM. rojo brillante (3B-A), rodamina y derivados: 6-carboxi-X-rodamina (ROX), 6-carboxirodamina (R6G), lisamina rodamina B, cloruro de sulfonilo rodamina (Rhod), rodamina B, rodamina 123, isotiocianato de rodamina X, sulforrodamina B, sulforrodamina 101, cloruro de sulfonilo derivado de la sulforrodamina 101 (Texas Red); N, N, N', N'tetrametil-6-carboxirodamina (TAMRA); tetrametilrodamina; isotiocianato de tetrametil-rodamina (TRITC); riboflavina; ácido rosólico; derivados de quelato de terbio; Cy3; Cy5; Cy5.5; Cy7; IRD 700; IRD 800; La Jolta Blue; ftalocianina; y naftalocianina. Los marcadores fluorescentes preferidos son cianina-3 y cianina-5. La invención contempla marcadores distintos de los marcadores fluorescentes, incluidos otros marcadores detectables ópticamente.

#### Polimerasas

Las polimerasas de ácidos nucleicos generalmente útiles en la invención incluyen polimerasas de ADN, polimerasas de ARN, transcriptasas inversas y formas mutantes o alteradas de cualquiera de las anteriores. Las ADN polimerasas y sus propiedades se describen en detalle, entre otros lugares, en DNA Replication, 2.ª edición, Kornberg y Baker, W. H. Freeman, Nueva York, NY. (1991). Las polimerasas de ADN convencionales conocidas útiles en la invención incluyen, aunque no de forma limitativa, la ADN polimerasa de *Pyrococcus furiosus* (Pfu) (Lundberg y col., 1991, Gene, 108: 1, Stratagene), *Pyrococcus woesei* (Pwo) DNA polymerase (Hinnisdaels y col., 1996, Biotechniques, 20:186-8, Boehringer Mannheim), *Thermus thermophilus* (Tth) DNA polymerase (Myers and Gelfand 1991, Biochemistry 30:7661), *Bacillus stearothermophilus* DNA polymerase (Stenesh and McGowan, 1977, Biochim Biophys Acta 475:32), *Thermococcus litoralis* (Tli) DNA polymerase (denominado también Vent.TM. DNA polymerase, Cariello y col., 1991, Polynucleotides Res, 19: 4193, New England Biolabs), 9.degree.Nm.TM. DNA polymerase (New England Biolabs), Stoffel fragment, ThermoSequenase® (Amersham Pharmacia Biotech UK), Therminator.TM. (New England Biolabs), *Thermotoga maritima* (Tma) DNA polymerase (Diaz and Sabino, 1998 Braz J. Med. Res, 31:1239), *Thermus aquaticus* (Taq) DNA polymerase (Chien y col., 1976, J. Bacteriol, 127: 1550), DNA polymerase, *Pyrococcus kodakaraensis* KOD DNA polymerase (Takagi y col., 1997, Appl. Environ. Microbial. 63:4504), JDF-3 DNA polymerase (from *thermococcus* sp. JDF-3, Patent application WO 0132887), *Pyrococcus* GB-D (PGB-D) DNA polymerase (also referred as Deep Vent.TM. DNA polymerase, Juncosa-Ginesta y col., 1994, Biotechniques, 16:820, New England Biolabs), UITma DNA polymerase (from thermophile *Thermotoga maritima*; Diaz and Sabino, 1998 Braz J. Med. Res, 31:1239; PE Applied Biosystems), Tgo DNA polymerase (from *thermococcus gorgonarius*, Roche Molecular Biochemicals), *E. coli* DNA polymerase I (Lecomte and Doubleday, 1983, Polynucleotides Res. 11:7505), la ADN polimerasa T7 (Nordstrom y col., 1981, J. Biol. Chem. 256:3112) y ADN-polimerasa II DP11/DP2 arqueal (Cann y col., 1998, Proc. Natl. Acad. Sci. USA 95:14250).

Se contemplan tanto las polimerasas mesófilas como las polimerasas termófilas. Las ADN-polimerasas termófilas incluyen, aunque no de forma limitativa, ThermoSequenase®, 9.Degree.NM.TM., Therminator.TM., Taq, Tne, Tma, Pfu, Tfi, Tth, Tli, Stoffel fragment, Vent.TM. y Deep Vent.TM. ADN polimerasa, ADN polimerasa KOD, Tgo, JDF-3 y mutantes, variantes y derivados de las mismas. Una forma altamente preferida de cualquier polimerasa es un mutante deficiente en exonucleasa 3'.

Las transcriptasas inversas útiles en la invención incluyen, aunque no de forma limitativa, las transcriptasas inversas del VIH, HTLV-1, HTLV-11, FeLV, FIV, SIV, AMY, MMTV, MoMuLV y otros retrovirus (véase Levin, Cell 88:5-8 (1997); Verma, Biochim Biophys Acta. 473:1-38 (1977); Wu y col., CRC Crit. Rev Biochem. 3:289-347 (1975)).

#### Adjunto

En una realización preferida, las moléculas molde de ácido nucleico se unen a un sustrato (también denominado en la presente memoria superficie) y se someten a análisis mediante secuenciación de una sola molécula como se describe en la presente memoria. Las moléculas molde de ácido nucleico se unen a la superficie de manera que los dúplex plantilla/cebador se pueden resolver ópticamente de forma individual. Los sustratos para su uso en la invención pueden ser bidimensionales o tridimensionales y pueden comprender una superficie plana (p. ej., un portaobjetos de vidrio) o pueden tener forma. Un sustrato puede incluir vidrio (p. ej., vidrio de poro controlado (CPG)), cuarzo, plástico (tal como poliestireno (poliestireno de baja reticulación y alta reticulación), policarbonato, polipropileno y poli (metacrilato de metilo)), copolímero acrílico, poliamida, silicio, metal (p. ej., oro derivatizado con alcanotiolato), celulosa, nailon, látex, dextrano, matriz de gel (p. ej., gel de sílice), poliácroleína o compuestos.

Los sustratos tridimensionales adecuados incluyen, p. ej., esferas, micropartículas, perlas, membranas, portaobjetos, placas, chips micromecanizados, tubos (p. ej., tubos capilares), micropocillos, dispositivos microfluidicos, canales, filtros o cualquier otra estructura adecuada para anclar un ácido nucleico. Los sustratos pueden incluir matrices o matrices planas capaces de tener regiones que incluyen poblaciones de ácidos nucleicos o cebadores modelo. Los ejemplos incluyen portaobjetos de poliestireno y CPG derivatizados con nucleósidos; diapositivas magnéticas derivatizadas; poliestireno injertado con polietilenglicol y similares.

Los sustratos se recubren preferiblemente para permitir un procesamiento óptico y una unión de ácidos nucleicos óptimos. Los sustratos para su uso en la invención también pueden tratarse para reducir el fondo. Los recubrimientos ilustrativos incluyen epóxidos y epóxidos derivatizados (p. ej., con una molécula de unión, tal como un oligonucleótido o estreptavidina).

Se pueden usar varios métodos para anclar o inmovilizar la molécula de ácido nucleico a la superficie del sustrato. La inmovilización se puede lograr mediante la unión directa o indirecta a la superficie. La unión puede ser por enlace covalente. Véase, Joos y col., *Analytical Biochemistry* 247:96-101, 1997; Oroskar y col., *Clin. Chem.* 42:1547-1555, 1996; y Khandjian, *Mol. Bio. Rep.* 11:107-115, 1986. Una unión preferida es la unión directa de amina de un nucleótido terminal de la plantilla o el extremo 5' del cebador a un epóxido integrado en la superficie. La unión también puede ser a través de un enlace no covalente. Por ejemplo, la biotina-estreptavidina (Taylor y col., *J. Phys. D. Appl. Phys.* 24:1443, 1991) y digoxigenina con anti-digoxigenina (Smith y col., *Science* 253: 1122, 1992) son herramientas comunes para anclar ácidos nucleicos a superficies y paralelos. Alternativamente, la unión se puede lograr anclando una cadena hidrófoba en una monocapa o bicapa lipídica. También se pueden usar otros métodos conocidos en la técnica para unir moléculas de ácido nucleico a sustratos.

#### Detección

Se puede usar cualquier método de detección que sea adecuado para el tipo de marcador empleado. Por lo tanto, los métodos de detección ilustrativos incluyen la detección radiactiva, la detección de absorbancia óptica, p. ej., la detección de absorbancia UV-visible, la detección de emisiones ópticas, p. ej., la fluorescencia o la quimioluminiscencia. Por ejemplo, los cebadores extendidos pueden detectarse en un sustrato escaneando todo o partes de cada sustrato simultáneamente o en serie, dependiendo del método de escaneo usado. Para el marcaje por fluorescencia, las regiones seleccionadas de un sustrato se pueden escanear en serie una por una o fila por fila usando un aparato de microscopio de fluorescencia, tal como se describe en Fodor (patente US-5.445.934) y Mathies y col. (Patente US-5.091.652). Los dispositivos capaces de detectar la fluorescencia de una sola molécula incluyen el microscopio de barrido de túnel (SiM) y el microscopio de fuerza atómica (AFM). Los patrones de hibridación también pueden escanearse usando una cámara CCD (p. ej., Model TE/CCD512SF, Princeton Instruments, Trenton, NJ.) con una óptica adecuada (Ploem, en *Fluorescent and Luminescent Probes for Biological Activity* Mason, T. G. Ed., Academic Press, Landon, págs. 1-11 (1993), tal como se describe en Yershov y col., *Proc. Natl. Acad. Sci.* 93:4913 (1996), o pueden ser captadas mediante un monitoreo de TV. Para señales radiactivas, se puede usar un dispositivo generador de imágenes de fósforo (Johnston y col., *Electrophoresis*, 13:566, 1990; Dermanac y col., *Electrophoresis*, 13:566, 1992; 1993). Otros proveedores comerciales de instrumentos de obtención de imágenes son General Scanning Inc., (Watertown, Massachusetts, en la World Wide Web en genscan.com), Genix Technologies (Waterloo, Ontario, Canadá); en la World Wide Web en confocal.com), y Applied Precision Inc. Estos métodos de detección son particularmente útiles para lograr el escaneo simultáneo de múltiples ácidos nucleicos de plantilla unidos.

Se pueden usar varios enfoques para detectar la incorporación de nucleótidos marcados con fluorescencia en una única molécula de ácido nucleico. Las configuraciones ópticas incluyen la microscopía de barrido de campo cercano, la microscopía confocal de campo lejano, la epiluminación de campo amplio, la dispersión de la luz, la microscopía de campo oscuro, la fotoconversión, la excitación de uno o varios fotones, la discriminación de longitudes de onda espectrales, la identificación de fluoróforos, la iluminación de ondas evanescentes y la microscopía de fluorescencia de reflexión interna total (TIRF). Generalmente, ciertos métodos implican la detección de la fluorescencia activada por láser utilizando un microscopio equipado con una cámara. Los sistemas de detección de fotones adecuados incluyen, aunque no de forma limitativa, fotodiodos y cámaras CCD intensificadas. Por ejemplo, se puede usar una cámara con dispositivo de acoplamiento de carga intensificada (ICCD). El uso de una cámara ICCD para obtener imágenes de moléculas de colorantes fluorescentes individuales en un fluido cerca de una superficie ofrece numerosas ventajas. Por ejemplo, con una configuración óptica ICCD, es posible adquirir una secuencia de imágenes (películas) de fluoróforos.

Algunas realizaciones de la presente invención usan microscopía TIRF para la obtención de imágenes. La microscopía TIRF utiliza luz de excitación totalmente reflejada internamente y es bien conocida en la técnica. Consulte, p. ej., la World Wide Web en nikon-instruments.jp/eng/page/products/tirf.aspx. En ciertas realizaciones, la detección se lleva a cabo usando iluminación de ondas evanescentes y microscopía de fluorescencia de reflexión interna total. Se puede establecer un campo de luz evanescente en la superficie, por ejemplo, para obtener imágenes de moléculas de ácido nucleico marcadas con fluorescencia. Cuando un rayo láser se refleja totalmente en la interfaz entre un sustrato líquido y sólido (p. ej., un vidrio), el haz de luz de excitación penetra solo una corta distancia en el líquido. El campo óptico no termina abruptamente en la interfaz reflectante, sino que su intensidad disminuye exponencialmente con la distancia. Este campo electromagnético superficial, llamado "onda evanescente", puede excitar selectivamente las moléculas fluorescentes del líquido cerca de la interfaz. El delgado campo óptico evanescente de la interfaz proporciona un fondo bajo y facilita la detección de moléculas individuales con una alta relación señal/ruido en longitudes de onda visibles.

El campo evanescente también puede obtener imágenes de nucleótidos marcados con fluorescencia tras su incorporación al complejo de plantilla/cebador adjunto en presencia de una polimerasa. La microscopía de fluorescencia de reflectancia interna total se usa luego para visualizar el dúplex de plantilla/cebador adjunto y/o los nucleótidos incorporados con una resolución de una sola molécula.

Algunas realizaciones de la invención utilizan métodos de detección no ópticos tales como, por ejemplo, la detección mediante nanoporos (p. ej., proteínas o estado sólido) a través de los cuales las moléculas pasan individualmente para permitir la identificación de las moléculas al observar las características o los cambios en diversas propiedades o efectos, tales como la capacitancia o el bloqueo del flujo de corriente (véase, por ejemplo, Stoddart y col., Proc. Nat. Acad. Sci., 106:7702, 2009; Purnell y Schmidt, ACS Nano, 3:2533, 2009; Branton y col., Nature Biotechnology, 26:1146, 2008; Polansky y col., solicitud de la patente US-2008/0187915; Mitchell y Howorka, Angew. Chem. Int. Ed. 47:5565, 2008; Borsenberger y col., J. Am. Chem. Soc., 131, 7530, 2009); u otros métodos de detección no ópticos adecuados.

#### Análisis

La alineación y/o compilación de los resultados de secuencia obtenidos a partir de las pilas de imágenes producidas como se ha descrito generalmente anteriormente utiliza tablas de consulta que tienen en cuenta los posibles cambios de secuencia (debidos, p. ej., a errores, mutaciones, etc.). Esencialmente, los resultados de secuenciación obtenidos como se describe en la presente memoria se comparan con una tabla de tipos de consulta que contiene todas las secuencias de referencia posibles más errores de 1 o 2 bases.

#### Determinación de la presencia de ácido nucleico fetal femenino en la muestra materna

Los métodos de la invención proporcionan un análisis cuantitativo o cualitativo adicional de los datos de secuencia para detectar la presencia de ácido nucleico fetal, independientemente de la capacidad de detectar el cromosoma Y, particularmente para detectar un feto femenino en una muestra materna. Generalmente, las secuencias obtenidas se alinean con un genoma de referencia (p. ej., un genoma materno, un genoma paterno o un estándar externo que representa el intervalo numérico considerado indicativo de una normalidad). Una vez alineadas, las secuencias obtenidas se cuantifican para determinar el número de lecturas de secuencia que se alinean con cada cromosoma. Se evalúan los recuentos cromosómicos y la desviación de una proporción normal de 2 veces proporciona evidencia de ácido nucleico fetal femenino en la muestra materna, y también proporciona evidencia de ácido nucleico fetal que representa aneuploidía cromosómica.

Se pueden realizar numerosos tipos diferentes de análisis cuantitativos para detectar la presencia de ácido nucleico fetal de un feto femenino en la muestra materna. Tal análisis adicional puede incluir análisis del número de copias, llamadas de alelos dispersos, resecuenciación dirigida, modificación diferencial del ADN (p. ej., metilación o bases modificadas) y análisis de puntos de ruptura. En ciertas realizaciones, no es necesario analizar los datos de secuencia para determinar la presencia de una parte del cromosoma Y, y los métodos de la invención pueden implicar la realización de un análisis cuantitativo como se describe en la presente memoria para detectar la presencia de ácido nucleico fetal en la muestra materna.

Un método para detectar la presencia de ácido nucleico fetal de un feto femenino en una muestra materna implica realizar un análisis del número de copias de los datos de secuencia generados. Este método implica determinar el cambio en el número de copias en los segmentos genómicos en relación con la información de la secuencia de referencia. La información de la secuencia de referencia puede ser una muestra materna que se sabe que no contiene ácido nucleico fetal (tal como una muestra bucal) o puede ser un estándar externo que represente el intervalo numérico considerado indicativo de un cariotipo normal e intacto. En este método, una cantidad enumerativa (número de copias) de un ácido nucleico diana (es decir, ADN cromosómico o una porción del mismo) en una muestra se compara con una cantidad enumerativa de un ácido nucleico de referencia. El número de referencia se determina por una cantidad estándar (es decir, esperada) del ácido nucleico en un cariotipo normal o por comparación con un número de un ácido nucleico de un cromosoma no objetivo en la misma muestra, conociéndose o sospechándose que el cromosoma no objetivo está presente en un número apropiado (es decir, diploide para los autosomas) en la muestra. La descripción adicional del análisis del número de copias se muestra en Lapidus y col. (patentes US-5.928.870 y US-6.100.029) y Shuber y col. (patente US-6.214.558).

El genoma humano normal contendrá solo números de copias integrales (p. ej., 0, 1, 2, 3, etc.), mientras que la presencia de ácido nucleico fetal en la muestra introducirá números de copias en valores fraccionarios (p. ej., 2,1). Si el análisis de los datos de la secuencia proporciona una colección de mediciones del número de copias que se desvían de los valores integrales esperados con significación estadística (es decir, superiores a los valores que se obtendrían debido a la varianza del muestreo, las inexactitudes de referencia o los errores de secuenciación), entonces la muestra materna contiene ácido nucleico fetal. Para una mayor sensibilidad, se puede usar una muestra de ácido nucleico materno y/o paterno para proporcionar información adicional sobre la secuencia de referencia. La información de secuencia de la muestra materna y/o paterna permite la identificación de los valores del número de copias en la muestra materna que se sospecha que contienen ácido nucleico fetal que no coinciden con la muestra de control materna y/o coinciden con la muestra paterna, que por lo tanto indica la presencia de ácido nucleico fetal. Otro método para detectar la presencia de ácido nucleico fetal de un feto femenino en una muestra materna implica realizar una llamada alélica dispersa. La llamada de alelos dispersos es un método que analiza alelos individuales en sitios polimórficos en una secuenciación de ADN de baja cobertura (p. ej., una cobertura inferior a 1x) para comparar las variaciones en los ácidos nucleicos de una muestra. El genoma de un individuo generalmente tiene alrededor de tres

mil millones de pares de bases de secuencia. Para un individuo típico, aproximadamente dos millones de posiciones son heterocigotas y aproximadamente un millón de posiciones son polimorfismos homocigotos de un solo nucleótido (SNP) no de referencia. Si se comparan dos mediciones de la misma posición alélica dentro de un individuo, coincidirán casi el 100 % de las veces en el caso de una posición homocigótica o casi el 50 % de las veces en el caso de una posición heterocigótica (los errores de secuenciación pueden disminuir ligeramente estos números). Si se comparan dos mediciones de la misma posición alélica en diferentes individuos, coincidirán con menos frecuencia, dependiendo de la frecuencia de los diferentes alelos en la población y de la relación entre los individuos. El grado de concordancia entre un amplio conjunto de posiciones alélicas en dos muestras es, por lo tanto, indicativo de la relación entre los individuos de los que se tomaron las muestras, donde cuanto más estrecha sea la relación, mayor será la concordancia (una muestra de un hermano o hijo, por ejemplo, será más similar a la muestra de un individuo que la de un extraño, pero menos similar que una segunda muestra del mismo individuo). La Figura 1 muestra los histogramas de la diferencia entre dos muestras de un individuo (“yo”) y muestras de ese individuo y dos miembros de la familia (“familia”) que representan la comparación de un conjunto de variantes conocidas de un solo nucleótido entre las diferentes muestras.

El método descrito anteriormente se puede utilizar para la detección de ADN fetal en una muestra materna comparando esta muestra con una muestra que incluye solo ADN materno (p. ej., una muestra bucal) y/o un ADN paterno. Este método implica obtener información de secuencia con una cobertura baja (p. ej., una cobertura inferior a 1x) para determinar si el ácido nucleico fetal está presente en la muestra. El método utiliza el hecho de que las variantes se producen en todo el genoma y millones de ellas están anotadas en bases de datos disponibles públicamente. La baja cobertura permite analizar un conjunto diferente de SNP en cada comparación. Se espera que la diferencia entre el genoma de un feto y el de su madre sea estadísticamente significativa si se buscan diferencias entre un número sustancial de las variantes que se encuentran en el genoma materno. Además, se espera que la similitud entre el genoma del feto y el ADN parental sea estadísticamente significativa, en comparación con una muestra materna pura, ya que el feto hereda la mitad de su ADN para su padre.

La invención implica comparar una secuencia de ADN genómico de baja cobertura (p. ej., una cobertura inferior a 1x) tanto de la muestra materna que se sospecha que contiene ADN fetal como de una muestra materna pura, en posiciones de variación de secuencia conocidas (de las bases de datos existentes) o sospechosas (a partir de los datos), y determinar si esa diferencia es mayor de lo que cabría esperar si dos muestras fueran ambas puramente maternas (es decir, no contenían ADN fetal). No se requiere una muestra del ADN paterno, pero podría usarse para aumentar la sensibilidad, donde la muestra paterna se compararía tanto con la muestra materna pura como con la muestra con ADN fetal sospechoso. Una mayor similitud estadísticamente significativa entre la muestra sospechosa y la muestra paterna sería indicativa de la presencia de ADN fetal.

Otro método para detectar la presencia de ácido nucleico fetal de un feto femenino en una muestra materna implica realizar una resecuenciación dirigida. La resecuenciación se muestra, por ejemplo, en Harris (solicitudes de las patentes US-2008/0233575, US-2009/0075252 y US-2009/0197257). En resumen, se selecciona un segmento específico de la diana (por ejemplo, mediante PCR, micromatrices o MIPS) antes de la secuenciación. Se introduce un cebador diseñado para hibridarse con este segmento particular y se forma un dúplex cebador/plantilla. El dúplex cebador/plantilla se expone a una polimerasa y al menos a un nucleótido marcado de forma detectable en condiciones suficientes para la adición de nucleótidos al cebador dependiente de la plantilla. Se determina la incorporación del nucleótido marcado, así como la identidad del nucleótido que es complementario a un nucleótido en la plantilla en una posición opuesta al nucleótido incorporado.

Después de la reacción de polimerización, el cebador puede retirarse del dúplex. La imprimación se puede eliminar por cualquier medio adecuado, por ejemplo, elevando la temperatura de la superficie o el sustrato de manera que el dúplex se derrita, o cambiando las condiciones del regulador para desestabilizar el dúplex, o una combinación de los mismos. Los métodos para fundir dúplex plantilla/cebador son bien conocidos en la técnica y se describen, por ejemplo, en el capítulo 10 de Molecular Cloning, a Laboratory Manual, 3.<sup>a</sup> edición suplementaria, J. Sambrook y D. W. Russell, Cold Spring Harbor Press (2001).

Después de retirar el cebador, el molde puede exponerse a un segundo cebador capaz de hibridarse con el molde. En un ejemplo, el segundo cebador es capaz de hibridarse con la misma región del molde que el primer cebador (también denominado en la presente memoria primera región), para formar un dúplex plantilla/cebador. La reacción de polimerización se repite a continuación, resecuenciando de este modo al menos una parte del molde.

La resecuenciación dirigida de regiones genómicas altamente variables permite una cobertura más amplia de esas regiones (p. ej., 1 Mb con una cobertura de 100 veces). Los genomas humanos normales contendrán variantes de un solo nucleótido a frecuencias de aproximadamente el 100 % o aproximadamente el 50 %, mientras que la presencia de ácido nucleico fetal introducirá posibles frecuencias adicionales (p. ej., el 10 %, el 60 %, el 90 %, etc.). Si el análisis de los datos de resecuencia proporciona una colección de frecuencias de variantes de secuencia que se desvían del 100 % o el 50 % con significación estadística (es decir, superiores a los valores que se obtendrían debido a la varianza del muestreo, las inexactitudes de referencia o los errores de secuenciación), entonces la muestra materna contiene ácido nucleico fetal.

Otro método para detectar la presencia de ácido nucleico fetal de un feto femenino en una muestra materna implica realizar un análisis que analice los puntos de ruptura. Un punto de ruptura de secuencia se refiere a un tipo de mutación que se encuentra en los ácidos nucleicos en la que secciones enteras de ADN se invierten, mezclan o reubican para crear nuevas uniones de secuencia que no existían en la secuencia original. Los puntos de interrupción de la secuencia pueden identificarse en la muestra materna sospechosa de contener ácido nucleico fetal y compararse con las muestras de control maternas y/o paternas. La aparición de un número estadísticamente significativo de puntos de ruptura identificados que no se detectan en la muestra de control materna y/o se detectan en la muestra paterna, indica la presencia de ácido nucleico fetal.

10 Detección de anomalías fetales

La capacidad de detectar el ácido nucleico fetal en una muestra materna permite el desarrollo de un ensayo de diagnóstico no invasivo para evaluar si el feto tiene una anomalía. Por lo tanto, otro aspecto de la invención proporciona métodos no invasivos que analizan el ácido nucleico fetal en una muestra materna para determinar si un feto tiene una anomalía. Los métodos de la invención implican obtener una muestra que incluye ácidos nucleicos maternos y fetales, realizar una reacción de secuenciación en la muestra para obtener información sobre la secuencia de los ácidos nucleicos de la muestra, comparar la información de secuencia obtenida con la información de secuencia de un genoma de referencia, determinando de este modo si el feto tiene una anomalía. En ciertas realizaciones, el genoma de referencia puede ser el genoma materno, el genoma paterno o una combinación de los mismos. En otros casos, el genoma de referencia puede ser un estándar externo que representa el intervalo numérico considerado indicativo de un cariotipo normal e intacto, tal como el genoma de referencia humano HG 18 actualmente existente.

Se pueden detectar diversas anomalías genéticas según los presentes métodos, incluida la aneuploidía (es decir, la aparición de uno o más cromosomas adicionales o ausentes) o alteraciones conocidas en uno o más genes, tales como el CFTR, el factor VIII (gen F8), la beta globina, la hemacromatosis, la G6PD, la neurofibromatosis, la GAPDH, la beta amiloide y la piruvato quinasa. Se conocen las secuencias y mutaciones comunes de esos genes. Se pueden detectar otras anomalías genéticas, tales como las que implican una secuencia que se elimina en un cromosoma humano, se mueve en una translocación o inversión, o se duplica en una duplicación cromosómica, en la que la secuencia se caracteriza por un trastorno genético conocido en el material genético fetal que no está presente en el material genético materno. Por ejemplo, las trisomías cromosómicas pueden incluir trisomías parciales, en mosaico, anulares, 18, 14, 13, 8, 6, 4, etc. Puede encontrar una lista de anomalías conocidas en el mapa mórbido de OMIM, <http://www.ncbi.nlm.nih.gov/Omim/getmorbidity.cgi>.

Estas anomalías genéticas incluyen mutaciones que pueden ser heterocigotas y homocigotas entre el ácido nucleico materno y fetal, y aneuploidías. Por ejemplo, una copia faltante del cromosoma X (monosomía X) produce el síndrome de Turner, mientras que una copia adicional del cromosoma 21 produce el síndrome de Down. Otras enfermedades, como el síndrome de Edward y el síndrome de Patau, son causadas por una copia adicional del cromosoma 18 y el cromosoma 13, respectivamente. El presente método puede usarse para detectar una translocación, adición, amplificación, transversión, inversión, aneuploidía, poliploidía, monosomía, trisomía, trisomía 21, trisomía 13, trisomía 14, trisomía 15, trisomía 16, trisomía 18, trisomía 22, triploidía, tetraploidía y anomalías de los cromosomas sexuales, que incluyen, pero no limitado a XO, XXY, XYY y XXX.

Los ejemplos de enfermedades en las que la secuencia diana puede existir en una copia del ADN materno (heterocigótica) pero causar enfermedad en un feto (homocigótica) incluyen la anemia drepanocítica, la fibrosis quística, la hemofilia y la enfermedad de Tay Sachs. En consecuencia, usando los métodos descritos aquí, se pueden distinguir los genomas con una mutación de los genomas con dos mutaciones.

La anemia falciforme es una enfermedad autosómica recesiva. El nueve por ciento de los afroamericanos estadounidenses son heterocigotos, mientras que el 0,2 % son homocigotos recesivos. El alelo recesivo provoca la sustitución de un solo aminoácido en las cadenas beta de la hemoglobina.

La enfermedad de Tay-Sachs es una enfermedad autosómica recesiva que provoca la degeneración del sistema nervioso. Los síntomas se manifiestan después del nacimiento. Los niños homocigotos recesivos para este alelo rara vez sobreviven más allá de los cinco años. Los pacientes carecen de la capacidad de producir la enzima N-acetilhexosaminidasa, que descompone el lípido gangliósido GM2.

Otro ejemplo es la fenilcetonuria (PKU), un trastorno hereditario recesivo en donde quienes la padecen carecen de la capacidad de sintetizar una enzima que convierta el aminoácido fenilalanina en tirosina. Las personas homocigotas recesivas para este alelo tienen una acumulación de fenilalanina y productos de degradación anormales en la orina y la sangre.

La hemofilia es un grupo de enfermedades en las que la sangre no se coagula normalmente. Los factores de la sangre están involucrados en la coagulación. Se considera que los hemofílicos que carecen del factor VIII normal tienen hemofilia A, y los que carecen del factor IX tienen hemofilia B. Estos genes se encuentran en el cromosoma X, por lo que los métodos de secuenciación de la invención se pueden usar para detectar si un feto heredó o no el cromosoma X defectuoso de la madre o el alelo normal del padre.

Se puede encontrar una lista de mutaciones genéticas para las que se pueden adaptar los presentes métodos en <http://www.gdb.org/gdb>, The GDB Human Genome Database, The Official Worldwide Database for the Annotation of the Human Genome Hosted by RTI International, Carolina del Norte, EE. UU.

5 Los cebadores específicos de los cromosomas se muestran en Hahn y col. (solicitud de la patente US-2005/0164241. Los cebadores para los genes se pueden preparar sobre la base de secuencias de nucleótidos obtenidas de bases de datos tales como GenBank, EMBL y similares. Por ejemplo, hay más de 1000 cebadores específicos del cromosoma 10 21 que figuran en el sitio web del UNIST del NIH, que se puede encontrar en <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>.

15 Un aspecto importante de un ensayo de diagnóstico es la capacidad del ensayo para distinguir entre falsos negativos (sin detección de ácido nucleico fetal) y negativos verdaderos (detección de ácido nucleico de un feto sano). Los métodos de la invención proporcionan esta capacidad detectando la presencia de al menos una porción de un cromosoma Y en la muestra, y también realizando un análisis adicional si el cromosoma Y no se detecta en la muestra. En ciertas realizaciones, los métodos de la invención distinguen entre falsos negativos y verdaderos negativos independientemente de la capacidad de detectar el cromosoma Y.

20 Si el cromosoma Y se detecta en la muestra materna, los métodos de la invención aseguran que el ensayo funcione correctamente, porque el cromosoma Y está asociado solo con los hombres y estará presente en una muestra materna solo si el ácido nucleico fetal masculino está presente en la muestra. Por lo tanto, si no se detecta ninguna anomalía en la muestra materna y se detecta al menos una parte del cromosoma Y en la muestra, se puede concluir con confianza que el ensayo ha detectado un feto (porque la presencia del cromosoma Y en una muestra materna es 25 indicativa del ácido nucleico fetal masculino) y que el feto no incluye la anomalía genética para la que se realizó el ensayo.

Los métodos de la invención también proporcionan un análisis cuantitativo o cualitativo adicional para detectar la presencia de ácido nucleico fetal independientemente de la capacidad de detectar el cromosoma Y. Esta etapa es particularmente útil en casos en los que la muestra incluye ácidos nucleicos normales de un feto femenino. Tal análisis 30 cuantitativo adicional puede incluir el análisis del número de copias, la llamada de alelos dispersos, la resecuenciación dirigida y el análisis de puntos de corte, cada uno de los cuales se discutió anteriormente. Por lo tanto, si no se detecta ninguna anomalía en la muestra materna y el análisis cuantitativo de la muestra revela la presencia de ácido nucleico fetal, se puede concluir con confianza que el ensayo ha detectado un feto y que el feto no incluye la anomalía genética para la que se realizó el ensayo.

#### 35 Etiquetado

En ciertos aspectos, un método de la invención determina si un feto tiene una anomalía obteniendo una muestra materna que incluye ácidos nucleicos maternos y fetales; unir marcadores únicos a los ácidos nucleicos de la muestra, 40 en las que cada marcador está asociado con un cromosoma diferente; realizar una reacción de secuenciación en los ácidos nucleicos marcados para obtener secuencias marcadas; y determinar si el feto tiene una anomalía cuantificando las secuencias marcadas.

45 La unión de marcadores a las secuencias diana se muestra en Kahvejian y col. (solicitud de la patente US-2008/0081330), y Steinman y col. (solicitud de la patente internacional PCT/US-09/64001). La secuencia de marcadora generalmente incluye ciertas características que hacen que la secuencia sea útil en las reacciones de secuenciación. Por ejemplo, los marcadores se diseñan para tener regiones homopolímeras mínimas o nulas, es decir, 2 o más de la misma base en una fila, tal como AA o CCC, dentro de la porción única del marcador. Los marcadores también se 50 diseñan para que estén al menos a una distancia de edición del orden de adición de bases al realizar la secuenciación base por base, garantizando que la primera y la última base no coincidan con las bases esperadas de la secuencia.

Los marcadores también pueden incluir bloqueadores, p. ej., nucleótidos que terminan la cadena, para bloquear la adición de bases al extremo 3' de las moléculas de ácido nucleico de la plantilla. Los marcadores también se diseñan para tener una similitud mínima con el orden de adición de bases, p. ej., si se realiza un método de secuenciación 55 base por base, generalmente las bases se agregan en el siguiente orden, una a la vez: C, T, A y G. Los marcadores también pueden incluir al menos un nucleótido no natural, tal como un ácido nucleico peptídico o un ácido nucleico bloqueado, para mejorar ciertas propiedades del oligonucleótido.

60 La parte de secuencia única del marcador (parte única) puede tener diferentes longitudes. Los métodos para diseñar conjuntos de marcadores únicos se muestran, por ejemplo, en Brenner y col. (patente US-6.235.475). En ciertas realizaciones, la parte única del marcador varía de aproximadamente 5 nucleótidos a aproximadamente 15 nucleótidos. En una realización particular, la parte única del marcador varía de aproximadamente 4 nucleótidos a aproximadamente 7 nucleótidos. Dado que la porción única del marcador se secuencia junto con la molécula de ácido nucleico molde, la longitud del oligonucleótido debe ser de longitud mínima para permitir la lectura más larga del ácido 65 nucleico molde adjunto. Generalmente, la porción única del marcador está separada de la molécula de ácido nucleico molde por al menos una base (minimiza las combinaciones homopoliméricas).

El marcador también incluye una parte que se usa como sitio de unión al cebador. El sitio de unión al cebador puede usarse para hibridar la molécula de ácido nucleico plantilla ahora codificada con código de barras con un cebador de secuenciación, que opcionalmente puede anclarse a un sustrato. La secuencia de unión al cebador puede ser una secuencia única que incluye al menos 2 bases, pero probablemente contiene un orden único de las 4 bases y generalmente tiene una longitud de 20-50 bases. En una realización particular, la secuencia de unión al cebador es un homopolímero de una sola base, p. ej. poliA, generalmente de 20 a 70 bases de longitud.

EL marcador también puede incluir un bloqueador, p. ej., un nucleótido que termina la cadena, en el extremo 3'. El bloqueador evita que se obtenga información de secuencia no deseada usando inadvertidamente el extremo 3' del sitio de unión del cebador como segundo cebador de secuenciación, particularmente cuando se usan secuencias de cebadores homopoliméricos. El bloqueador puede ser cualquier resto que impida que una polimerasa añada bases durante la incubación con un dNTP. Un bloqueador ilustrativo es un terminador nucleotídico que carece de un 3'-OH, es decir, un didesoxinucleótido (ddNTP). Los terminadores nucleotídicos comunes son 2',3'-didesoxinucleótidos, 3'-aminonucleótidos, 3'-desoxinucleótidos, 3'-azidonucleótidos, aciclonucleótidos, etc. El bloqueador puede tener adherido un marcador detectable, p. ej., un fluoróforo. El marcador se puede unir mediante un enlace lábil, p. ej., un disulfuro, de modo que, tras la hibridación del ácido nucleico molde con código de barras con la superficie, las ubicaciones de los ácidos nucleicos molde se puedan identificar mediante imágenes. Generalmente, el marcador detectable se retira antes de comenzar con la secuenciación. Dependiendo del enlace, el producto escindido puede requerir o no una modificación química adicional para evitar reacciones secundarias indeseables, por ejemplo, después de la escisión de un disulfuro por el TCEP, el tiol reactivo producido se bloquea con yodoacetamida.

Los métodos de la invención implican unir el marcador a las moléculas de ácido nucleico de la plantilla. Los ácidos nucleicos de plantilla pueden fragmentarse o cortarse a la longitud deseada, p. ej., generalmente de 100 a 500 bases o más, usando una variedad de métodos mecánicos, químicos y/o enzimáticos. El ADN puede cortarse aleatoriamente mediante sonicación, p. ej., el método de Covaris, una exposición breve a una ADNasa o el uso de una mezcla de una o más enzimas de restricción, o una transposasa o enzima melladora. El ARN puede fragmentarse por exposición breve a una RNasa, calor más magnesio, o por cizallamiento. El ARN se puede convertir en ADNc antes o después de la fragmentación.

En ciertas realizaciones, el marcador se une a la molécula de ácido nucleico molde con una enzima. La enzima puede ser una ligasa o una polimerasa. La ligasa puede ser cualquier enzima capaz de ligar un oligonucleótido (ARN o ADN) a la molécula de ácido nucleico molde. Las ligasas adecuadas incluyen la ligasa de ADN T4 y la ligasa de ARN T4 (tales ligasas están disponibles comercialmente en New England Biolabs). En una realización particular. Los métodos para usar ligasas son bien conocidos en la técnica. La polimerasa puede ser cualquier enzima capaz de añadir nucleótidos al extremo 3' de las moléculas de ácido nucleico molde. La polimerasa puede ser, por ejemplo, la poli (A) polimerasa de levadura, disponible comercialmente en USB. La polimerasa se usa según las instrucciones del fabricante.

La ligadura puede tener un extremo romo o mediante el uso de extremos colgantes complementarios. En ciertas realizaciones, después de la fragmentación, los extremos de los fragmentos pueden repararse, recortarse (p. ej., usando una exonucleasa) o rellenarse (p. ej., usando una polimerasa y dNTP), para formar extremos romos. Tras la generación de extremos romos, los extremos pueden tratarse con una polimerasa y dATP para formar una adición independiente del molde al extremo 3' de los fragmentos, produciendo por lo tanto una sola A que sobresale. Esta única A se usa para guiar la ligadura de fragmentos con una sola T que sobresale del extremo 5' en un método denominado clonación T-A.

Alternativamente, dado que la posible combinación de salientes dejados por las enzimas de restricción es conocida después de una digestión de restricción, los extremos pueden dejarse tal cual, es decir, extremos irregulares. En ciertas realizaciones, se usan oligonucleótidos bicatenarios con extremos colgantes complementarios. En un ejemplo particular, se usa el método A:T de sobrecolgarse con una sola base (véanse las Figuras 1-2).

En una realización particular, el sustrato ha anclado un complemento inverso a la secuencia de unión al cebador del oligonucleótido, por ejemplo, 5'-TC CAC TTA TCC TTG CAT CCA TCC TCT GCC CTG o un poliT(50). Cuando se usan secuencias homopoliméricas para el cebador, puede ser ventajoso realizar un procedimiento conocido en la técnica como "rellenar y bloquear". Cuando el poliA (20-70) de la muestra y el poliT (50) de la superficie se hibridan, existe una alta probabilidad de que no haya una alineación perfecta, por lo que el híbrido se completa incubando la muestra con polimerasa y TTP. Tras la etapa de llenado, la muestra se lava y la polimerasa se incuba con uno o dos dNTP complementarios a la(s) base(s) utilizada(s) en la secuencia de bloqueo. El llenado y el bloqueo también se pueden realizar en un proceso de una sola etapa en donde la polimerasa, el TTP y uno o dos terminadores reversibles (complementos de las bases de bloqueo) se mezclan y se incuban. Los terminadores reversibles detienen la adición durante esta etapa y pueden volver a ser funcionales (inversión del mecanismo inhibidor) mediante tratamientos específicos para los análogos utilizados. Algunos terminadores reversibles tienen bloqueos funcionales en el 3'-OH que deben eliminarse, mientras que otros, por ejemplo, los terminadores virtuales Helicos BioSciences, tienen inhibidores unidos a la base mediante un disulfuro que se puede eliminar mediante el tratamiento con TCEP.

Una vez marcados, los ácidos nucleicos de la muestra materna se secuencian como se describe en la presente memoria. Los marcadores permiten que los ácidos nucleicos molde de diferentes cromosomas se diferencien entre sí a lo largo del proceso de secuenciación. Debido a que cada uno de los marcadores está asociado con un cromosoma diferente, las secuencias marcadas se pueden cuantificar. Las lecturas de secuencia se evalúan para detectar cualquier desviación con respecto a una proporción normal de 2 veces, desviación que indica una anomalía fetal.

En una alternativa, el ácido nucleico materno libre de células recibe un código de barras antes de la secuenciación ligando las secuencias de códigos de barras al extremo 3' de los fragmentos de ADN materno. Un código de barras preferido es de 5 a 8 nucleótidos, que se usan como identificadores únicos del ADN libre de células maternas. Esas secuencias también pueden incluir una cola de polinucleótido de 50 nt (p. ej., poli-A). Hacer esto permite la posterior hibridación del ácido nucleico directamente con la superficie de la célula de flujo seguida de la secuenciación. Entre otras cosas, este método permite la combinación de diferentes muestras de ADN materno en un único canal celular de flujo para la secuenciación, que permite por lo tanto multiplexar las reacciones.

#### 15 Detección de secuencias únicas

En ciertos aspectos, los métodos de la descripción se usan para detectar el ácido nucleico fetal obteniendo una muestra materna que se sospecha que incluye ácido nucleico fetal, detectando al menos dos secuencias únicas en la muestra y determinando si el ácido nucleico fetal está presente en la muestra materna basándose en la proporción de las secuencias detectadas entre sí. Las secuencias únicas son secuencias que se sabe que aparecen solo una vez en el genoma relevante (p. ej., humano) y pueden conocerse como k-meros únicos o pueden determinarse mediante secuenciación. De forma ventajosa, estos métodos de la invención no requieren comparación con una secuencia de referencia. En una muestra materna, cabría esperar que se produjeran dos o más k-mer únicos con una frecuencia idéntica, lo que lleva a una proporción de 1,0. Una varianza estadísticamente significativa con respecto a la ración esperada es indicativa de la presencia de ácido nucleico fetal en la muestra.

En ciertas realizaciones, una o más secuencias k-mer únicas están predeterminadas basándose en el conocimiento disponible de los k-meros únicos en el genoma humano. Por ejemplo, es posible estimar el número de k-meros únicos en cualquier genoma basándose en la secuencia de consenso. El conocimiento de la presencia real de secuencias únicas de cualquier número dado de bases está fácilmente disponible para los expertos en la materia relevante.

En un ejemplo, se hace un recuento del número de veces que se detectan dos o más secuencias únicas en la muestra materna. Por ejemplo, la secuencia A (p. ej., una secuencia única de 20 meros) puede detectarse 80 veces y la secuencia B (p. ej., una secuencia de 30 meros única) puede detectarse 100 veces. Si la secuencia se detecta uniformemente en todo el genoma humano, o al menos para las porciones que incluyen las secuencias A y B, entonces el ácido nucleico fetal que tiene la secuencia B está presente en la muestra materna a un nivel por encima del origen materno indicado al menos en parte por la proporción de (100 -80) a 80. En la medida en que la secuencia no se detecte de manera uniforme, se pueden emplear varios métodos conocidos de análisis estadístico para determinar si la diferencia medida entre la frecuencia de la secuencia A y la secuencia B es estadísticamente significativa.

Además, se puede seleccionar cualquiera de las secuencias A, B o ambas para que tengan un contenido (p. ej., rico en GC) de modo que la detección uniforme se base más probablemente en factores conocidos por los expertos en la materia. Se puede seleccionar un gran número de secuencias únicas para hacer que la comparación estadística sea más sólida. Además, las secuencias se pueden seleccionar con respecto a su ubicación en una región genómica de particular interés. Por ejemplo, las secuencias pueden seleccionarse debido a su presencia en un cromosoma asociado con la aneuploidía. Por lo tanto, en ciertas realizaciones, si la secuencia A (detectada 80 veces) se hubiera seleccionado con respecto a su ubicación no en un cromosoma asociado con la aneuploidía, y la secuencia B (detectada 100 veces) se hubiera seleccionado con respecto a su ubicación dentro de un cromosoma asociado con la aneuploidía, podría hacerse un diagnóstico de aneuploidía fetal.

En otras realizaciones, las secuencias únicas incluyen uno o más SNP conocidos en ubicaciones conocidas. Además de contar el número de veces que se detecta la secuencia A en la muestra materna, también se puede contar el número de veces que la secuencia A tiene una variante en una ubicación del SNP conocida (por ejemplo, una "G") y el número de veces que la secuencia A tiene la otra variante en esa ubicación del SNP (p. ej., una "T"). Mientras la madre y el feto no sean homocigotos para la misma base en esa ubicación, la señal fetal puede detectarse mediante cualquier desviación de G o T con respecto a los niveles estadísticamente probables (hasta cualquier nivel de certeza deseado) suponiendo cualquier otra combinación de cigosidad. Para el caso en donde tanto la madre como el feto sean homocigotos en la ubicación del SNP, se puede realizar una comparación con otra o más secuencias únicas predeterminadas (como la secuencia B) como se describió anteriormente.

En otro enfoque más, las secuencias detectadas no necesitan ser únicas y no necesitan estar predeterminadas. Además, no es necesario saber nada sobre el genoma humano (u otro). Más bien, la firma de la madre puede distinguirse de la firma del feto (si está presente) basándose en un patrón de n-meros (o n-meros y k-meros, etc.). Por ejemplo, en cualquier patrón de n-meros, habrá SNP, de modo que la madre tenga una base (p. ej., "G") y el feto, si está presente, tenga otra base (p. ej., "T") en al menos uno de los dos alelos. Si todos los n-meros lf (en una muestra suficientemente grande en vista de cualquier tasa de error) tienen una "G", entonces se puede decir que no hay ácido

nucleico fetal. Si algún número estadísticamente significativo de n-meros tiene una “T” en la ubicación del SNP, entonces se ha detectado el ácido nucleico fetal y se puede determinar la cantidad, en relación con el ácido nucleico de la madre. Esto es cierto aunque puede haber dos o más lugares en los que el n-mero aparezca en uno o ambos genomas de la madre o del feto (es decir, las secuencias no son únicas), porque, dado un número suficientemente grande de lecturas, habrá una diferencia estadísticamente significativa en los SNP detectados con respecto a la presencia o falta de señal fetal. Es decir, habrá una diferencia estadísticamente significativa en la frecuencia de los alelos que se detectan entre lo que cabría esperar de un solo organismo contribuyente en lugar de dos (o más).

### Ejemplos

#### Ejemplo 1: Determinación de la presencia de ácido nucleico fetal en una muestra

Se obtuvieron muestras de ácido nucleico de linfocitos de hombres y mujeres adultos sanos normales. Los ácidos nucleicos se extrajeron mediante protocolos conocidos en la técnica. El conjunto de muestras incluía 2 tríos HapMap (6 muestras) ejecutados en 8 canales del secuenciador HELISCOPE (instrumento de secuenciación de una sola molécula, Helicos BioSciences Corporation) en 3 máquinas diferentes (2 réplicas técnicas). El ADN genómico de una de las muestras se secuenció en cada canal (8-13 millones de lecturas alineadas de forma única).

El conjunto de datos incluye 8 archivos comprimidos, uno para cada canal de HELISCOPE. Las lecturas de secuencia se mapearon en un genoma humano de referencia y se descartaron las lecturas con alineaciones no únicas (Figura 2). Los recuentos se normalizaron primero por muestra, basándose en los recuentos totales de los cromosomas autosómicos (Figura 3). A continuación, los recuentos se normalizaron por cromosoma, basándose en la fracción promedio de lecturas alineadas con cada cromosoma en todas las muestras (chrX, solo mujeres, chrY, solo hombres); Figura 4).

Los datos muestran un análisis cromosómico cuantitativo (Figura 5). Estos datos muestran la secuenciación genómica de muestras seleccionadas de HapMap, tanto masculinas como femeninas, seguida de una cuantificación precisa de los recuentos cromosómicos. Los datos de la presente memoria muestran la capacidad distintiva para identificar las proporciones esperadas del cromosoma X y el cromosoma Y. Los datos derivados del ADN genómico obtenido de individuos demuestran la uniformidad de la cobertura genómica esperada de un genoma diploide normal y demuestran que no se encuentra ácido nucleico fetal en estas muestras. La desviación en los recuentos normalizados por cromosoma es de 0,5 % de CV en promedio. Es menor (0,2-0,3 %) para los cromosomas más grandes y mayor (0,8-1,1 %) para los cromosomas más pequeños. Las muestras femeninas y masculinas se distinguen claramente.

#### Ejemplo 2: Detección del ácido nucleico fetal en una muestra materna y detección de la trisomía

El ácido nucleico plasmático libre de células maternas se obtuvo usando métodos muy conocidos en la técnica, tales como un kit de purificación de ácidos nucleicos Qiagen. El ácido nucleico se sometió después al siguiente protocolo. En resumen, el protocolo consiste en una etapa posterior de poliA 3' de una hora, seguida de una etapa de bloqueo con didesoxilo de una hora en 3'. El protocolo se llevó a cabo con 500 pg de ácido nucleico.

#### Reactivos necesarios

Kit de transferasa terminal	NEB M0315
dATP	Roche 11277049001
Biotina-ddATP	Perkin Elmer NEL548001
Oligonucleótido portador	Oligonucleótido de 50 meros
Albúmina sérica bovina	NEB B9001S
Agua libre de nucleasas	
Reactivo Quant-iT™ PicoGreen dsDNA	Invitrogen PI 1495

#### Equipo necesario

Bloque de aluminio preenfriado fresado para tubos de 0,2 ml

Termociclador

P-2, P20, P200

Cubo de hielo con pipeta

Nanodrop 3300 o un lector de placas estándar para el ensayo de PicoGreen

## Métodos

5 Antes de llevar a cabo la reacción residual en el ADN, se eliminó la contaminación por ARN mediante la digestión con RNasa y la limpieza con un kit de limpieza de reacciones de Qiagen (catálogo 28204). El ADN se cuantificó con precisión antes de su uso. Se usó el kit de reactivos de Quant-iT™ PicoGreen dsDNA (Invitrogen, catálogo n.º PI 1495) con un fluoroespectrómetro Nanodrop 3300. Se usó glucógeno libre de nucleasas de grado biológico molecular o acrilamida lineal como portador durante las etapas de limpieza/precipitación del ADN.

15 Se preparó la siguiente mezcla: Regulador NEB de transferasa terminal 10 veces (2 µl); 2,5 mM de CoCH (2 µl); y ácido nucleico plasmático exento de células maternas y agua sin nucleasas (10,8 µl). El volumen total fue de 14,8 µl. La mezcla se calentó a 95 °C durante 5 minutos en el termociclador para desnaturalizar el ADN. Después del calentamiento, la mezcla se enfrió en el bloque de aluminio preenfriado que se mantuvo en una suspensión de hielo y agua (aproximadamente 0 °C) para obtener ADN monocatenario. La muestra se enfrió lo más rápido posible para evitar la rehibridación del ADN monocatenario desnaturalizado.

20 En hielo, se añadió la siguiente mezcla al ADN desnaturalizado de arriba: 1 µl de transferasa terminal (diluir de 1:4 a 5 U/µl con regulador Ix); 4 µl de dATP 50 µM; y 0,2 µl de BSA. El volumen de esta mezcla fue de 5,2 µl, lo que llevó el volumen total de la reacción a 20 µl. Los tubos que contenían la mezcla se colocaron en el termociclador y se ejecutó el siguiente programa: 37 °C durante 1 hora; 70 °C durante 10 minutos; y la temperatura se redujo de nuevo a 4 °C. Ahora se habrá agregado una cola de poli (A) al ADN.

25 La reacción de poliadenilación de 20 µl se desnaturalizó calentando la mezcla a 95 °C durante 5 minutos en el termociclador, seguido de un enfriamiento rápido en el bloque de aluminio preenfriado mantenido en una suspensión de hielo y agua (aproximadamente 0 °C). La muestra se enfrió lo más rápido posible para evitar la rehibridación del ADN monocatenario desnaturalizado.

30 La siguiente mezcla de bloqueo se añadió a la mezcla poliadenilada desnaturalizada desde arriba: 1 µl de regulador de transferasa terminal 10 veces; 1 µl de CoCH (2,5 mM); 1 µl de transferasa terminal (diluir de 1:4 a 5 U/µl con regulador Ix); 0,5 µl de biotina-DDATP 200 µM; y 6,5 µl de agua libre de nucleasas. El volumen de esta mezcla fue de 10 µl, lo que llevó el volumen total de la reacción a 30 µl.

35 Los tubos que contenían la mezcla se colocaron en el termociclador y se ejecutó el siguiente programa: 37 °C durante 1 hora; 70 °C durante 20 minutos; y la temperatura se redujo de nuevo a 4 °C. Se observó que ahora se añadía un bloque terminal 3' al ADN poliadenilado.

40 Se añadieron 2 picomoles de oligonucleótido de control a la reacción de transferasa terminal de 30 µl inactivada por calor anterior. El oligonucleótido de control se añadió a la muestra para minimizar la pérdida de ADN durante las etapas de carga de la muestra. El oligonucleótido de control no contiene una cola de poli (A) y, por lo tanto, no se hibridará con la superficie de la célula de flujo. La muestra ahora está lista para hibridarse con las celdas de flujo para la reacción de secuenciación. No se requiere ninguna etapa de limpieza adicional.

45 Las muestras se cargaron en los canales del secuenciador HELISCOPE (instrumento de secuenciación de una sola molécula, Helicos BioSciences Corporation) según las instrucciones del fabricante. El ADN de la muestra se secuenció en los canales según las instrucciones del fabricante. Las lecturas de secuencia se mapearon en un genoma humano de referencia y se descartaron las lecturas con alineaciones no únicas. Los recuentos se normalizaron primero por muestra, basándose en los recuentos totales de los cromosomas autosómicos. Luego, los recuentos se normalizaron por cromosoma, basándose en la fracción promedio de lecturas alineadas con cada cromosoma en todas las muestras (ChRx: solo mujeres, ChRy, solo hombres). Los recuentos cromosómicos de los cromosomas 1, 18 y 21 en las muestras se compararon con las desviaciones de los valores esperados según las muestras de control.

55 La Figura 10 muestra los resultados del análisis de la información de secuencia. En esta figura, el cromosoma 1 se usó como control. Los datos de la presente memoria muestran que se detectó ADN fetal (Figura 10). Los datos de la presente memoria muestran además que también se detectó la trisomía del cromosoma 18 y el cromosoma 21 (Figura 10).

### Ejemplo 3: Corregir el sesgo de GC

60 Cuando se realiza un análisis de recuento cromosómico con respecto a la información de secuenciación (es decir, cuantificando la cantidad de cada cromosoma o segmento cromosómico, con respecto a la representación relativa), se compara un número relativo de recuentos de lectura de cada cromosoma (o segmento cromosómico) con un estándar medido en una o más muestras normales. Ciertas etapas del proceso de preparación o secuenciación de la muestra pueden provocar un sesgo de GC, en donde la representación relativa de cada cromosoma está influenciada no solo por la cantidad relativa (número de copias) de ese cromosoma, sino también por su contenido de GC. Una diferencia en el sesgo de GC entre la muestra medida y la muestra de control (normal) provocará un

sesgo en los recuentos cromosómicos, de modo que los cromosomas con un contenido extremo de GC pueden parecer tener más o menos copias que su número real. La Figura 6 es un gráfico que muestra una muestra en la que los recuentos cromosómicos están sesgados por el sesgo de GC. Los cromosomas se ordenan aumentando el contenido de GC. Estos datos muestran que la variabilidad de la medición es mayor para los cromosomas con un contenido extremo de GC.

Los métodos de la invención permiten determinar una cantidad de sesgo de GC en la información de secuencia obtenida, y también permiten la corrección del sesgo de GC en la información de secuencia. Los métodos de la invención implican secuenciar una muestra para obtener información sobre la secuencia de ácidos nucleicos; determinar una cantidad de sesgo de GC en la información de secuencia; corregir la información de secuencia para tener en cuenta el sesgo de GC; y analizar la información corregida.

La determinación de la cantidad de sesgo de GC en una muestra se puede lograr de numerosas maneras. En ciertas realizaciones, la cantidad de sesgo de GC puede cuantificarse dividiendo el genoma en grupos y midiendo la correlación entre el número de recuentos en cada bin y su contenido de GC. La Figura 7 es un gráfico que muestra los recuentos en cada bin trazados con respecto al contenido de GC del bin. En esta realización, el genoma se divide en bins de 1000 kpb. Aunque este número es ilustrativo y se puede usar cualquier tamaño. Una correlación negativa o positiva significativa indica la existencia de un sesgo de GC (véase Figura 7). En la Figura 7, la muestra superior muestra una correlación positiva con el contenido de GC, y la muestra inferior muestra una correlación negativa con el contenido de GC.

Los métodos de la invención reducen o eliminan los efectos del sesgo de GC en la información de secuencia. Se pueden usar numerosos protocolos para reducir o eliminar los efectos del sesgo de GC en la información de secuencia. En ciertas realizaciones, se selecciona un subconjunto de bins genómicos dentro de un intervalo dado, de modo que el contenido promedio de GC por cromosoma se iguala (o menos sesgado). El recuento cromosómico se realiza entonces en el subconjunto seleccionado. La Figura 8 proporciona un ejemplo de este protocolo. En la Figura 8, el análisis se limitó solo a los bins genómicos con un contenido de GC dado de 0,42 a 0,48, aproximadamente el 25 % del genoma (Figura 8, panel A).

Los paneles B y C de la Figura 8 muestran la diferencia en la información de secuencia obtenida después de que haya una corrección para el sesgo de GC en la información de secuencia. El panel B de la Figura 8 muestra la información de la secuencia antes de la corrección del sesgo de GC. El panel C de la Figura 8 muestra la información de la secuencia después de la corrección del sesgo de GC. Estos datos muestran que el sesgo de GC estaba sesgando los recuentos cromosómicos, de modo que los cromosomas con un contenido extremo de GC parecían tener más o menos copias que su número real. Después de la corrección del sesgo de GC en la información de la secuencia, los datos muestran un recuento cromosómico más preciso y permiten la detección de la trisomía en los cromosomas 18 y 21, lo que no era posible a partir del análisis de la información de la secuencia antes de la corrección del sesgo de GC.

En otros casos, que no se reivindican, la correlación entre el contenido de GC y los recuentos de cromosomas se modela en un conjunto de bins genómicos utilizando una función matemática (p. ej., un polinomio de primer o segundo orden). Una función matemática ilustrativa es un modelo de regresión (es decir, ajustar la información de secuencia a una función matemática, tal como funciones de orden inferior (polinomios lineales y/o cuadráticos)). El efecto del sesgo de GC se corrige restando el componente dependiente del control de calidad, reflejado por el modelo, del recuento de cada bin. A continuación, se realiza el recuento cromosómico con respecto a los recuentos corregidos. Una ventaja de esta realización es que conserva el número de recuentos del conjunto de datos original, lo cual es importante para la sensibilidad del método.

La Figura 9 proporciona un ejemplo de este protocolo. En la Figura 9, la información de la secuencia se corrigió restando un modelo lineal de dependencia del GC de cada bin genómico. Los paneles A y B de la Figura 9 muestran información de secuencia antes de la corrección del sesgo de GC. Los paneles C y D de la Figura 9 muestran información de secuencia después de la corrección del sesgo de GC. Estos datos muestran que el sesgo de GC estaba sesgando los recuentos cromosómicos, de modo que los cromosomas con un contenido extremo de GC parecían tener más o menos copias que su número real. Después de la corrección del sesgo de GC en la información de la secuencia, los datos muestran un recuento cromosómico más preciso y permiten la detección de la trisomía en los cromosomas 18 y 21, lo que no era posible a partir del análisis de la información de la secuencia antes de la corrección del sesgo de GC.

En otras realizaciones adicionales, que no se reivindican, el sesgo de GC se corrige de la siguiente manera. Se obtiene una cobertura promedio por bin en varias muestras de control y la cobertura observada en la muestra se divide por la media de la población de control (esta podría ser una media ponderada para tener en cuenta los diferentes niveles de cobertura general en las muestras de control). Cada valor de bin corregido sería entonces una relación entre lo observado y lo esperado, que será más uniforme en todos los bins de diferentes porcentajes de GC.

**REIVINDICACIONES**

1. Un método para el recuento cromosómico con efectos reducidos o eliminados del sesgo de GC en la información de secuencia, comprendiendo el método:
 

5                                    secuenciar una muestra para obtener información sobre la secuencia de ácidos nucleicos; corregir la información de secuencia para tener en cuenta el sesgo de GC seleccionando un subconjunto de bins dentro de un intervalo dado de valores de contenido de GC de modo que se iguale el contenido promedio de GC por cromosoma; y

10                                    analizar la información de secuencia corregida realizando un recuento cromosómico utilizando los recuentos del subconjunto seleccionado de bins.
2. El método según la reivindicación 1, que comprende:
 

15                                    comparar la información de secuencia corregida con una secuencia de referencia; identificar el ácido nucleico fetal en la muestra; y determinar si el feto tiene una anomalía basándose en la comparación con la secuencia de referencia.
3. El método según la reivindicación 1 o 2, en donde los bins del subconjunto de bins seleccionado tienen un intervalo de contenido de GC de aproximadamente 0,42 a 0,48.
4. El método según la reivindicación 1 o 2, en donde el subconjunto seleccionado de bins representa aproximadamente el 25 % de un genoma de referencia.
5. El método según una cualquiera de las reivindicaciones 1 a 4, en donde el tamaño de un bin es de aproximadamente 1000 pares de kilobases (kb).
6. El método según una cualquiera de las reivindicaciones 1 a 5, en donde la muestra es un tejido o fluido corporal.
7. El método según la reivindicación 6, en donde se sospecha que la muestra contiene ácido nucleico fetal.
8. El método según la reivindicación 7, en donde el fluido corporal es sangre materna, plasma sanguíneo o suero.
9. El método según una cualquiera de las reivindicaciones 1 a 8, en donde la secuenciación es la secuenciación de una sola molécula, la secuenciación por síntesis y/o la secuenciación por detección de nanoporos.
10. El método según la reivindicación 2, en donde la secuencia de referencia se selecciona entre una secuencia de referencia materna, una secuencia de referencia fetal o una secuencia genómica humana consensuada.
11. El método según la reivindicación 10, en donde dicha secuencia de referencia materna se selecciona entre una secuencia obtenida de una muestra bucal, una muestra de saliva, una muestra de orina, una muestra de aspirado del pezón mamario, una muestra de esputo, una muestra de lágrima y una muestra de líquido amniótico.
12. El método según la reivindicación 2, en donde el ácido nucleico fetal es un ácido nucleico fetal circulante libre de células.
13. El método según la reivindicación 2, en donde antes de la etapa de secuenciación, el método comprende además el enriquecimiento del ácido nucleico fetal en la muestra.
14. El método según la reivindicación 2, en donde la etapa de identificación comprende una técnica seleccionada entre la llamada de alelos dispersos, la secuenciación génica dirigida, la identificación del material cromosómico Y, la enumeración, el análisis del número de copias y el análisis de inversión.

Figura 1

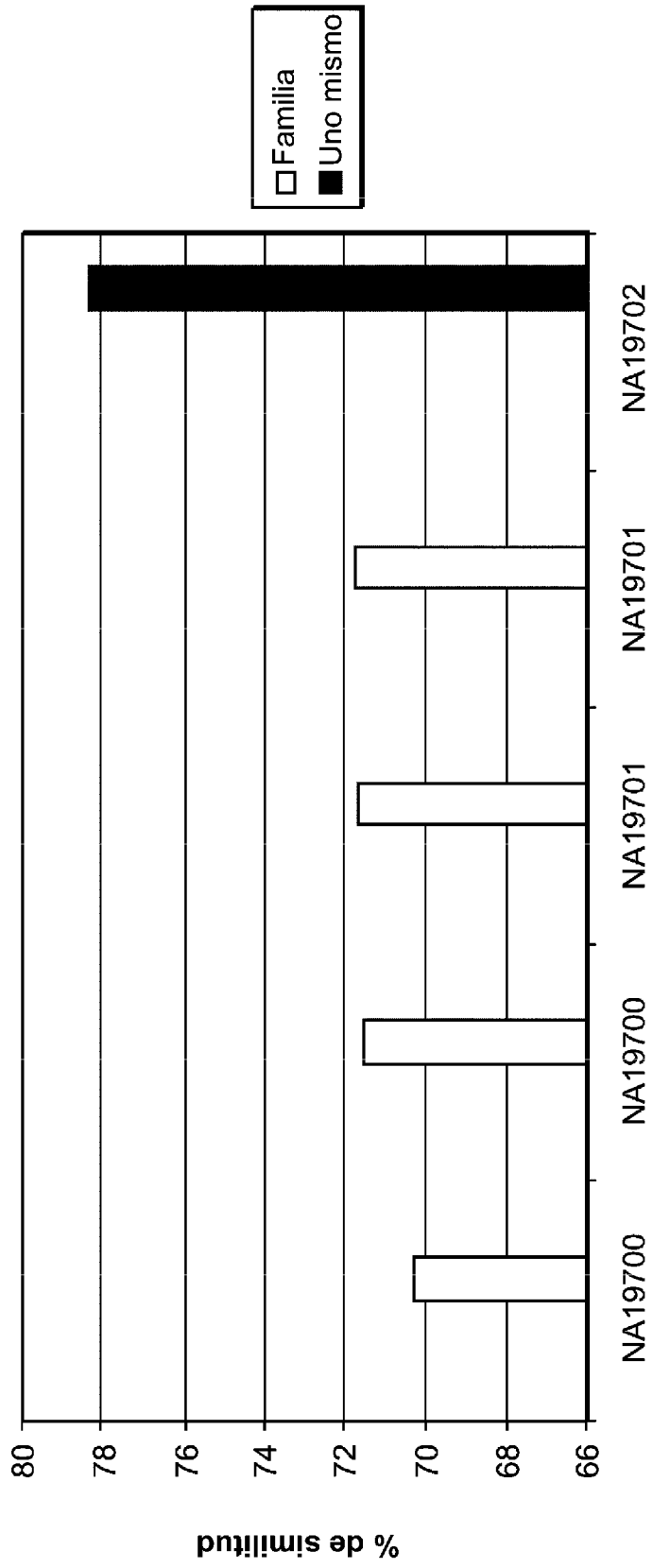


Figura 2

Muestra M/F Ref	NA19701 F CTS4_090409.fc1.ch14	NA19908 M CTS4_090409.fc1.ch15	NA19700 M CTS7_090109.fc1.ch23	NA19702 M CTS7_090109.fc1.ch24
<b>Recuentos sin procesar</b>				
chr1	755.789	949.344	762.507	707.587
chr2	810.736	1.015.940	820.013	757.528
chr3	670.166	839.989	678.781	626.727
chr4	645.458	809.796	655.001	602.821
chr5	601.593	756.338	609.740	562.562
chr6	581.171	727.129	583.912	541.122
chr7	513.499	646.457	518.952	481.056
chr8	494.837	619.535	499.760	461.678
chr9	372.595	468.371	375.386	346.900
chr10	455.424	567.192	453.109	421.747
chr11	448.070	561.852	449.132	416.253
chr12	446.393	558.399	450.471	418.417
chr13	329.439	413.156	335.090	307.840
chr14	298.056	373.264	301.173	278.891
chr15	268.724	338.013	268.179	250.797
chr16	260.075	324.986	255.926	239.438
chr17	246.242	307.373	244.640	228.824
chr18	261.478	329.137	264.399	243.182
chr19	171.905	216.479	171.028	160.990
chr20	204.819	256.213	203.060	188.511
chr21	120.474	152.048	121.702	112.275
chr22	108.757	135.416	106.386	99.598
chrX	459.668	319.889	258.971	241.708
chrY	4.840	31.316	25.573	23.710
<b>Total (autosoma)</b>	<b>9.065.700</b>	<b>11.366.427</b>	<b>9.128.347</b>	<b>8.454.744</b>

Figura 2 (continuación)

NA19909 F P5_082109.fc2.ch12	NA19919 M P5_082109.fc2.ch13	NA19700 M P5_082109.fc2.ch8	NA19701 F P5_082109.fc2.ch9
722.941	777.001	953.359	974.424
768.077	829.823	1.020.821	1.040.881
636.806	688.794	843.403	863.635
608.057	656.168	812.077	829.216
570.772	617.889	755.993	774.642
546.335	589.423	724.272	743.948
488.165	525.887	646.214	659.560
471.748	508.848	623.274	634.989
355.376	383.626	469.629	479.410
428.299	464.280	567.586	583.315
424.492	459.765	561.469	573.515
425.346	458.383	562.444	575.032
310.306	335.126	413.660	422.851
284.537	307.896	375.924	383.347
251.946	272.025	336.037	345.625
245.348	266.147	321.064	332.907
235.921	254.370	307.813	316.506
245.956	265.919	328.433	335.409
166.635	177.609	216.438	221.927
195.050	211.063	256.461	261.812
112.838	123.678	151.979	155.237
102.990	111.989	134.874	137.409
441.147	262.590	322.593	600.427
5.474	25.908	31.512	6.866
<b>8.597.941</b>	<b>9.285.709</b>	<b>11.383.224</b>	<b>11.645.597</b>

Figura 3

Normalizado por canal	Ref	CTS4_090409.fc1.ch14	CTS4_090409.fc1.ch15	CTS7_090109.fc1.ch23	CTS7_090109.fc1.ch24	P5_082109.fc2.ch12
chr1		0,083367969	0,083521761	0,083531772	0,083691121	0,08408304
chr2		0,089428946	0,08938077	0,089831489	0,089597982	0,08933267
chr3		0,073923249	0,073900884	0,074359684	0,074127259	0,07406494
chr4		0,071197812	0,071244552	0,071754612	0,071299734	0,07072123
chr5		0,066359244	0,066541403	0,066796321	0,066538029	0,06638473
chr6		0,064106578	0,063971642	0,063966893	0,064002174	0,06354254
chr7		0,056641958	0,056874249	0,0568506	0,056897761	0,05677697
chr8		0,05458343	0,054505695	0,054748138	0,054605793	0,05486755
chr9		0,041099419	0,041206529	0,041123108	0,041030219	0,04133269
chr10		0,050235944	0,049900642	0,049637574	0,049882882	0,04981414
chr11		0,049424755	0,049430837	0,049201898	0,04923307	0,04937136
chr12		0,049239772	0,049127048	0,049348584	0,049489021	0,04947068
chr13		0,036339058	0,036348802	0,036708727	0,036410328	0,03609073
chr14		0,032877329	0,032839167	0,032993159	0,032986333	0,03309362
chr15		0,029641837	0,029737841	0,029378704	0,029663465	0,02930306
chr16		0,028687801	0,028591747	0,028036401	0,028319959	0,02853567
chr17		0,02716194	0,027042183	0,026800033	0,027064569	0,02743924
chr18		0,02884256	0,028956945	0,028964609	0,028762787	0,02860638
chr19		0,018962132	0,019045475	0,018735922	0,019041381	0,0193808
chr20		0,02259274	0,022541208	0,022244991	0,022296476	0,02268566
chr21		0,01328899	0,013376939	0,013332315	0,013279527	0,01312384
chr22		0,011996536	0,011913682	0,011654465	0,011780132	0,01197845
chrX		0,050704082	0,028143321	0,028369978	0,028588447	0,05130845
chrY		0,00053388	0,002755131	0,002801493	0,002804343	0,00063666

Figura 3 (continuación)

P5_082109.fc2.ch13	P5_082109.fc2.ch8	P5_082109.fc2.ch9	Promedio
0,0836771	0,0837512	0,0836732	<b>0,083662</b>
0,0893656	0,0896777	0,0893798	<b>0,089499</b>
0,0741779	0,0740918	0,0741598	<b>0,074101</b>
0,0706643	0,0713398	0,0712042	<b>0,071178</b>
0,0665419	0,0664129	0,066518	<b>0,066512</b>
0,0634764	0,0636263	0,0638823	<b>0,063822</b>
0,056634	0,056769	0,056636	<b>0,05676</b>
0,054799	0,0547537	0,0545261	<b>0,054674</b>
0,0413136	0,0412562	0,0411666	<b>0,041191</b>
0,0499994	0,0498616	0,0500889	<b>0,049928</b>
0,0495132	0,0493243	0,0492474	<b>0,049343</b>
0,0493644	0,0494099	0,0493776	<b>0,049353</b>
0,0360905	0,0363394	0,0363099	<b>0,03633</b>
0,033158	0,0330244	0,0329178	<b>0,032986</b>
0,029295	0,0295204	0,0296786	<b>0,029527</b>
0,028662	0,028205	0,0285865	<b>0,028453</b>
0,0273937	0,0270409	0,0271782	<b>0,02714</b>
0,0286374	0,0288524	0,0288014	<b>0,028803</b>
0,0191271	0,0190138	0,0190567	<b>0,019045</b>
0,0227299	0,0225297	0,0224816	<b>0,022513</b>
0,0133192	0,0133511	0,0133301	<b>0,0133</b>
0,0120604	0,0118485	0,0117992	<b>0,011879</b>
0,0282789	0,0283393	0,0515583	<b>0,05119</b> (Solo mujeres)
0,0027901	0,0027683	0,0005896	<b>0,002784</b> (Solo hombres)

Figura 4

Ref	CTS4_090409.fc1.ch14	CTS4_090409.fc1.ch15	CTS7_090109.fc1.ch23	CTS7_090109.fc1.ch24	P5_082109.fc2.ch12
chr1	1,00	1,00	1,00	1,00	1,01
chr2	1,00	1,00	1,00	1,00	1,00
chr3	1,00	1,00	1,00	1,00	1,00
chr4	1,00	1,00	1,01	1,00	0,99
chr5	1,00	1,00	1,00	1,00	1,00
chr6	1,00	1,00	1,00	1,00	1,00
chr7	1,00	1,00	1,00	1,00	1,00
chr8	1,00	1,00	1,00	1,00	1,00
chr9	1,00	1,00	1,00	1,00	1,00
chr10	1,01	1,00	0,99	1,00	1,00
chr11	1,00	1,00	1,00	1,00	1,00
chr12	1,00	1,00	1,00	1,00	1,00
chr13	1,00	1,00	1,01	1,00	0,99
chr14	1,00	1,00	1,00	1,00	1,00
chr15	1,00	1,01	0,99	1,00	0,99
chr16	1,01	1,00	0,99	1,00	1,00
chr17	1,00	1,00	0,99	1,00	1,01
chr18	1,00	1,01	1,01	1,00	0,99
chr19	1,00	1,00	0,98	1,00	1,02
chr20	1,00	1,00	0,99	0,99	1,01
chr21	1,00	1,01	1,00	1,00	0,99
chr22	1,01	1,00	0,98	0,99	1,01
chrX	0,99	0,55	0,55	0,56	1,00
chrY	0,19	0,99	1,01	1,01	0,23

Figura 4 (continuación)

P5_082109.fc2.ch13	P5_082109.fc2.ch8	P5_082109.fc2.ch9	CV
1,00	1,00	1,00	0,3%
1,00	1,00	1,00	0,2%
1,00	1,00	1,00	0,2%
0,99	1,00	1,00	0,5%
1,00	1,00	1,00	0,2%
0,99	1,00	1,00	0,4%
1,00	1,00	1,00	0,2%
1,00	1,00	1,00	0,2%
1,00	1,00	1,00	0,3%
1,00	1,00	1,00	0,4%
1,00	1,00	1,00	0,2%
1,00	1,00	1,00	0,2%
0,99	1,00	1,00	0,5%
1,01	1,00	1,00	0,3%
0,99	1,00	1,01	0,6%
1,01	0,99	1,00	0,8%
1,01	1,00	1,00	0,8%
0,99	1,00	1,00	0,5%
1,00	1,00	1,00	0,9%
1,01	1,00	1,00	0,8%
1,00	1,00	1,00	0,6%
1,02	1,00	0,99	1,1%
0,55	0,55	1,01	
1,00	0,99	0,21	

Figura 5

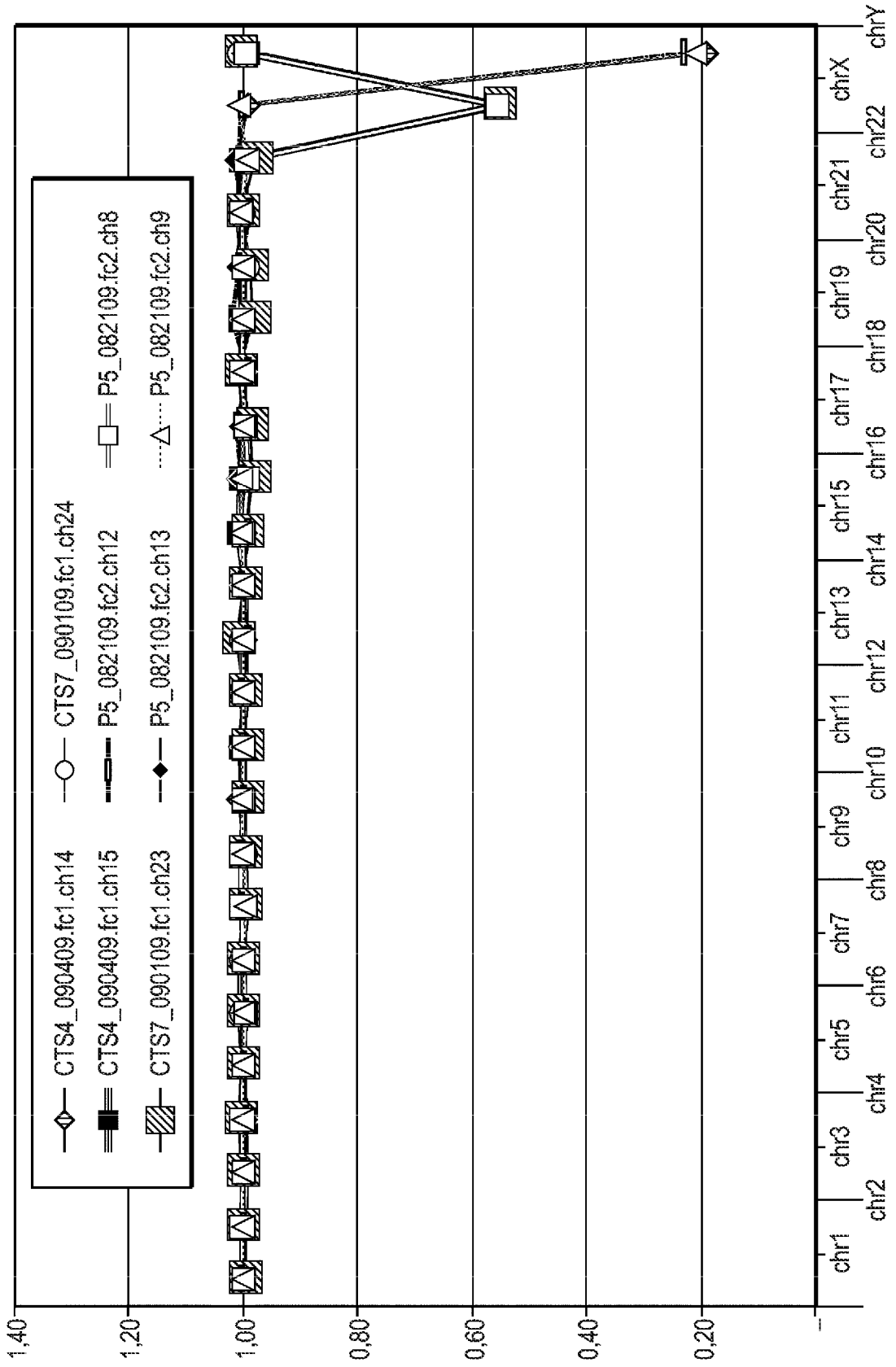


Figura 6A

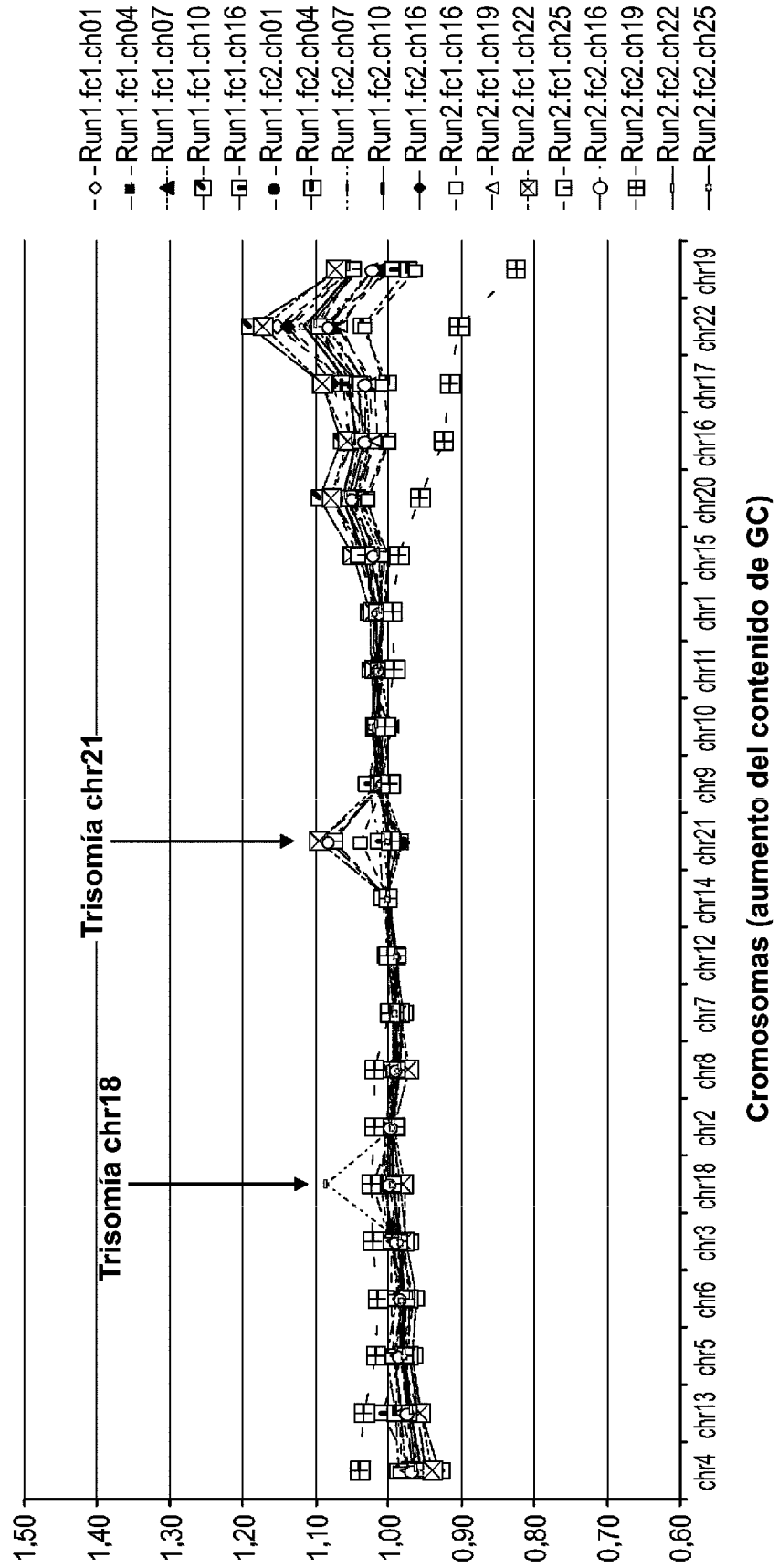


Figura 6B

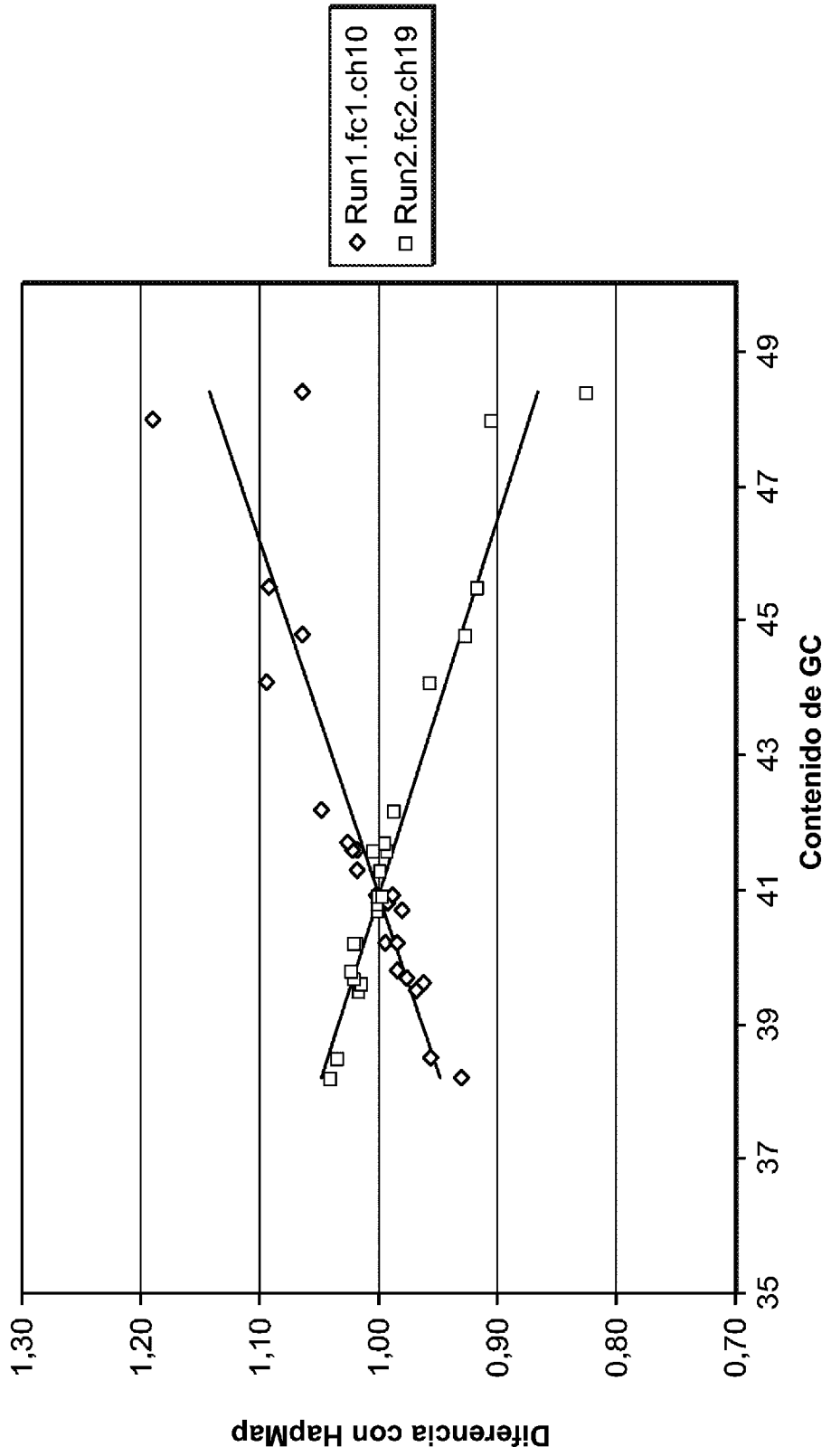


Figura 7

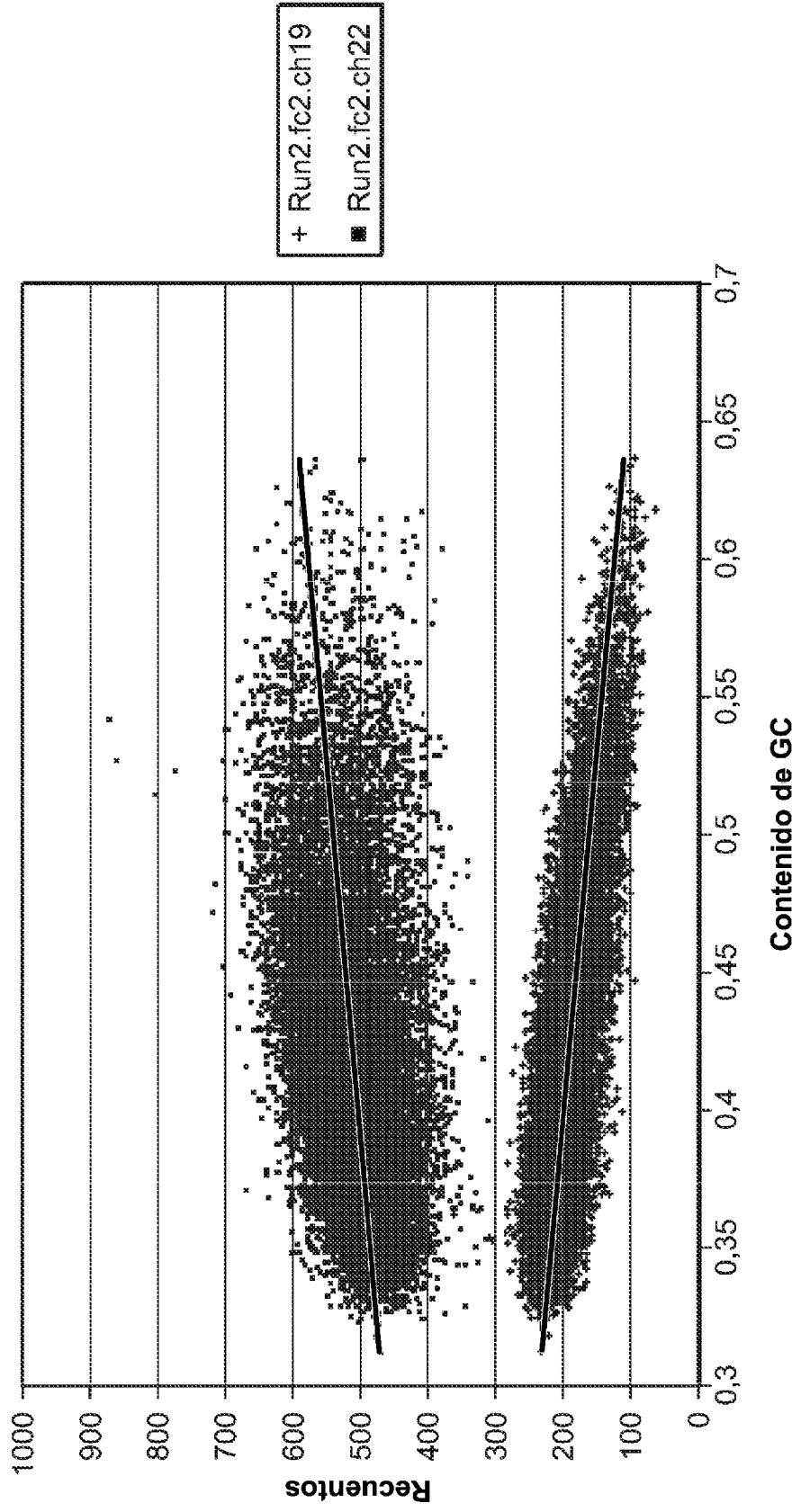


Figura 8A

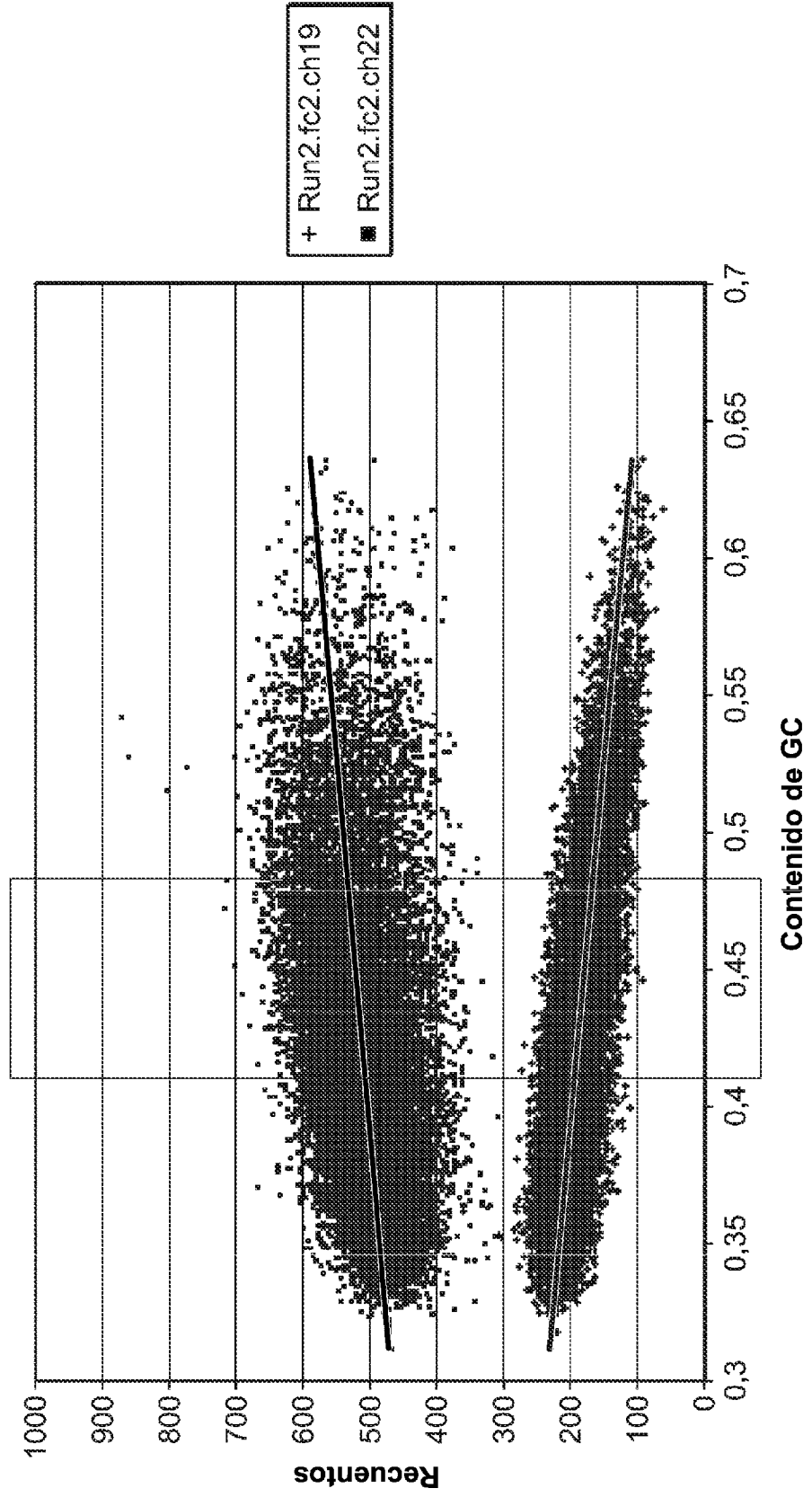


Figura 8B

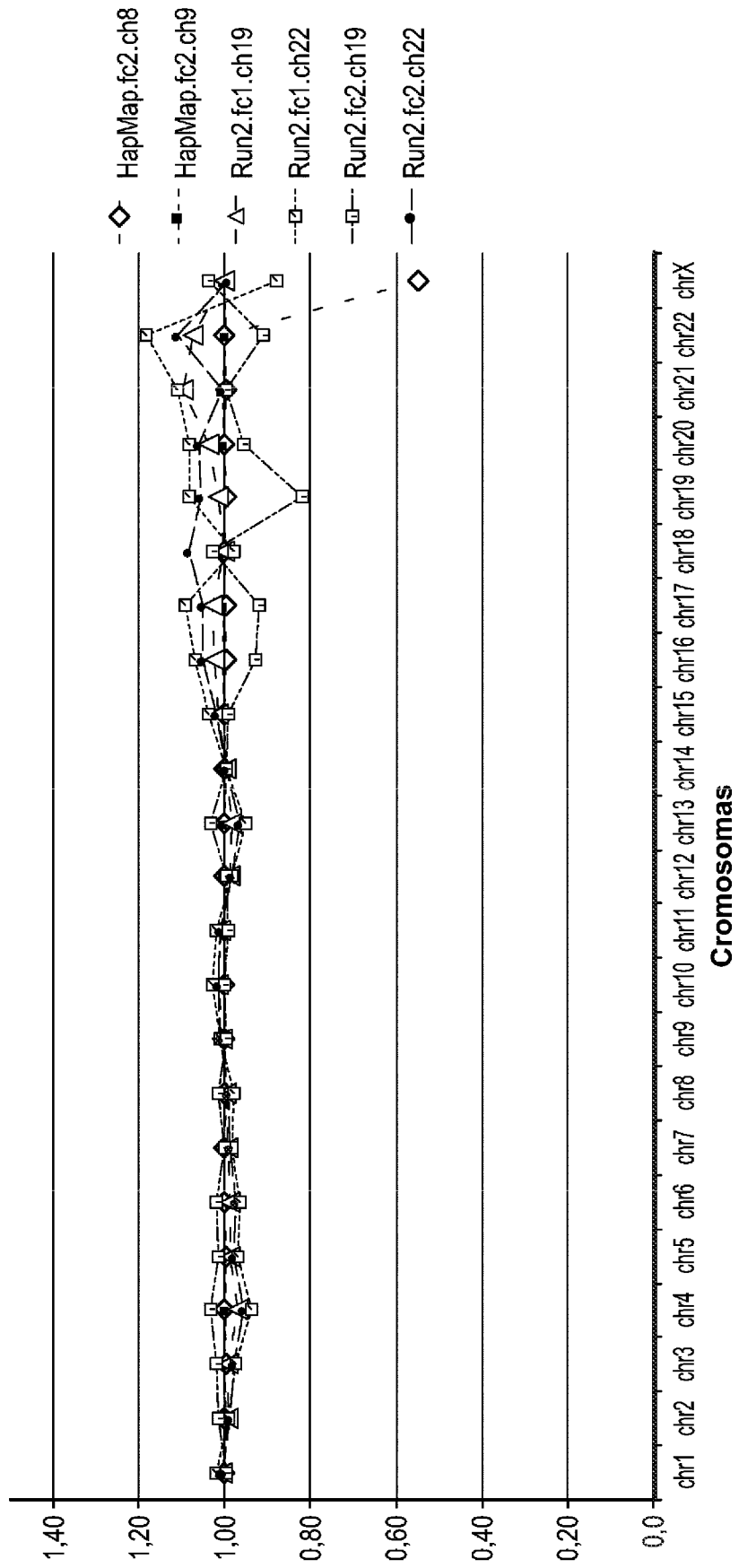


Figura 8C

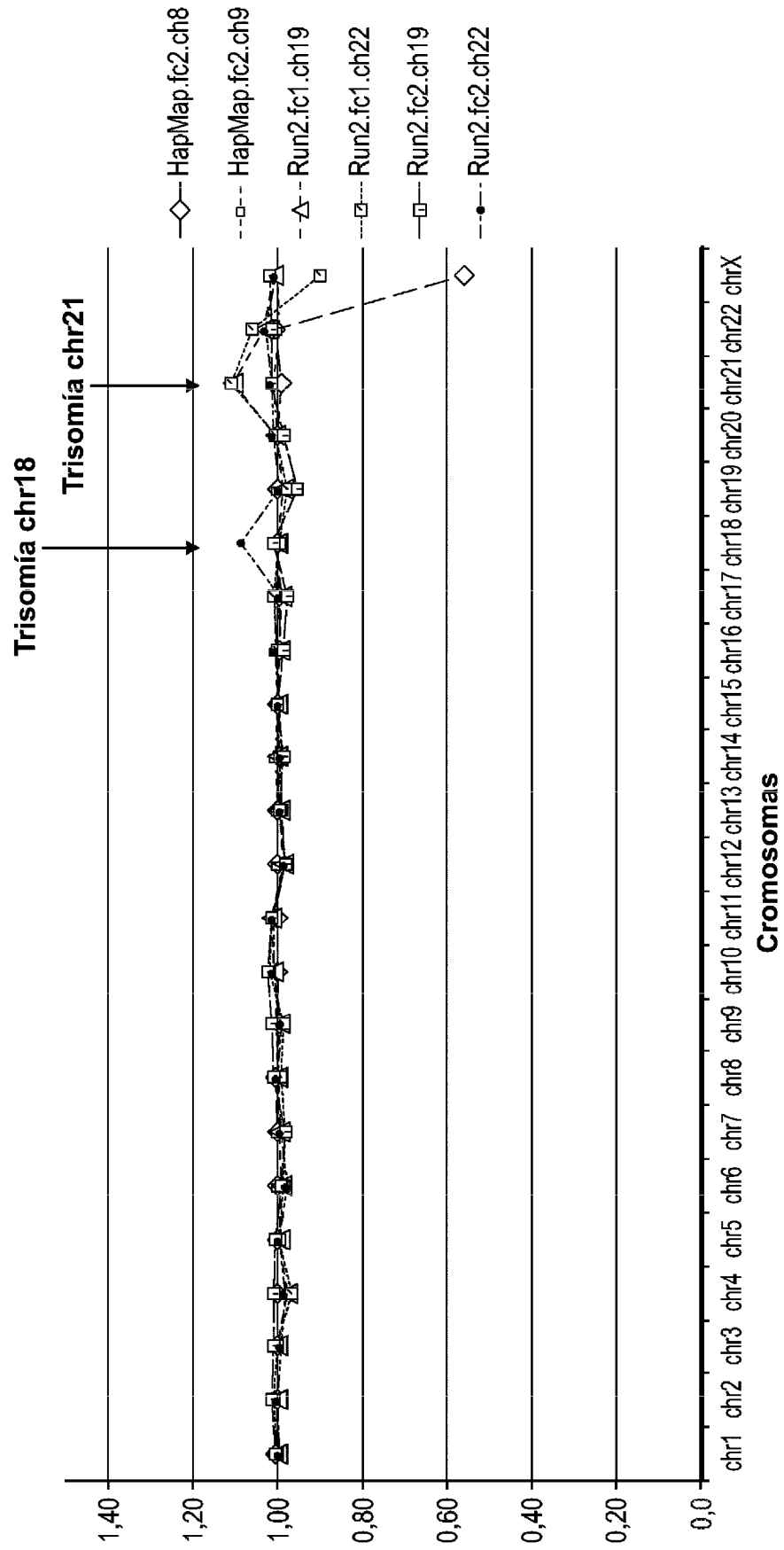


Figura 9A

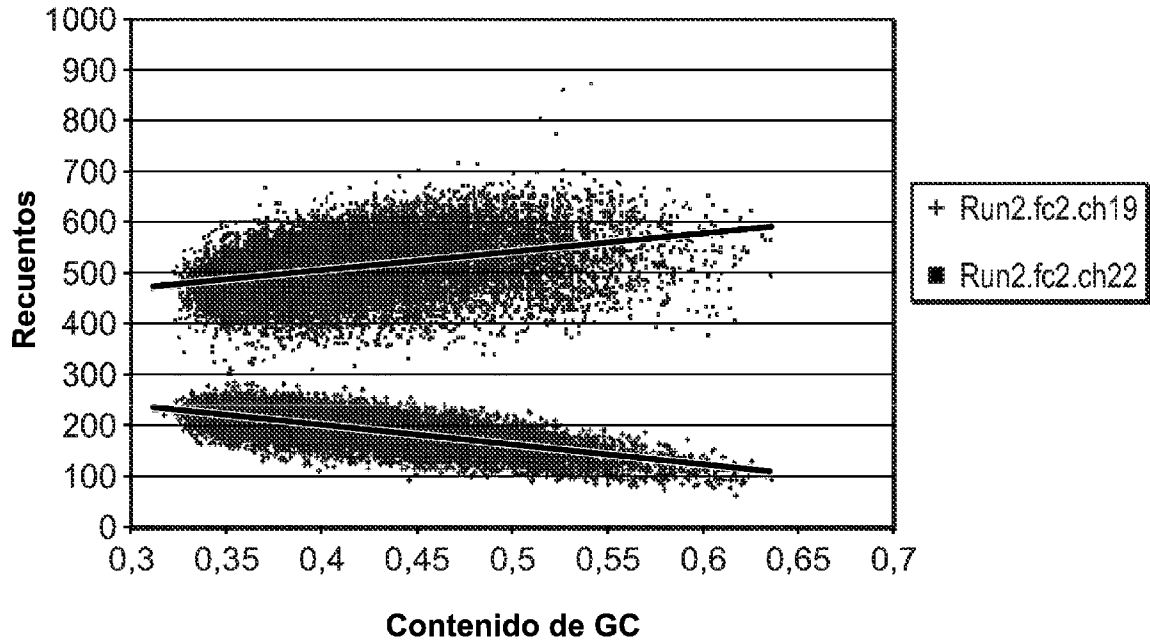


Figura 9C

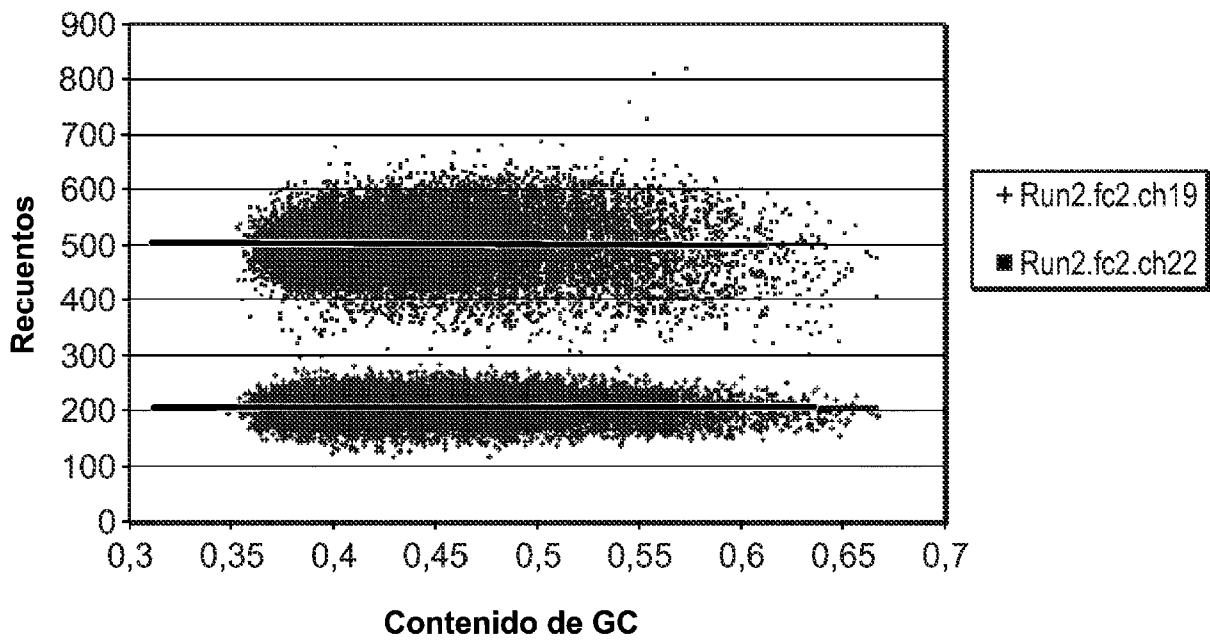


Figura 9B

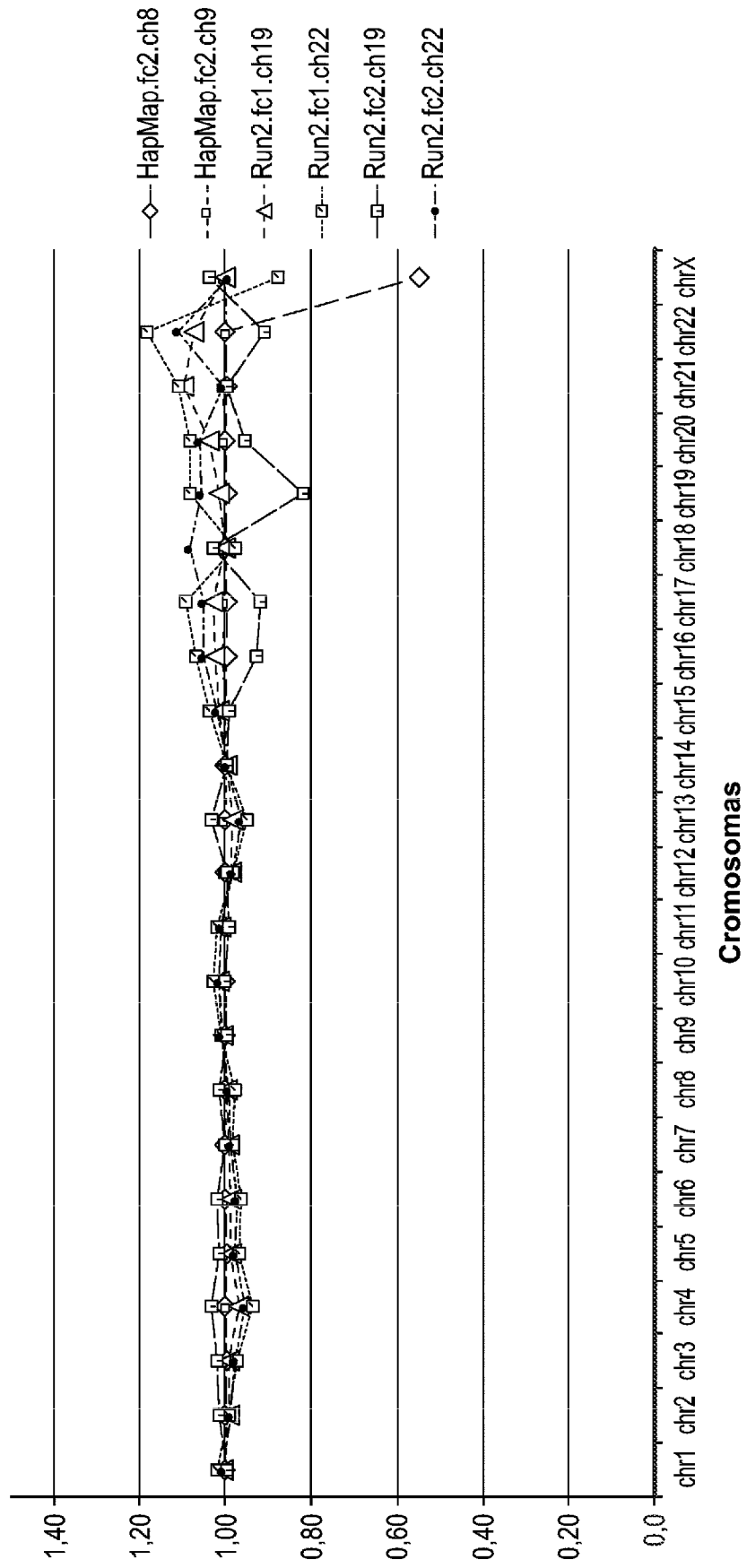


Figura 9D

